湖岸
Hu'an

Douglas Hofstadter

[美] 侯世达 著
修佳明 译



I AM A
STRANGE
LOOP

我是个怪圈

最终，我们自我感知、发明自我，并被锁入一场叫作“我”的幻景，
是自我指涉的小小奇迹。

中信出版集团

版权信息

书名:我是个怪圈

作者:[美]侯世达

译者:修佳明

ISBN:9787508695792

中信出版集团制作发行

"I", the balance between self-knowledge and self-ignorance.

"我"，自知与无知的平衡。

——侯世达

谨以此书献给能理解我的姐姐劳拉
同时献给不能理解我的妹妹茉莉

致

丽兹·马圭尔

1958—2006

她永远活在我们心中

出版方附记

作为基本图书公司多年来的老朋友，侯世达特意将本页留给我们，以纪念一位已故的编辑。

致谢

心灵是什么？心灵做什么？从10多岁开始，我就被这些谜题深深吸引，并潜心思索了几十年。我得出的结论，有一部分是来自个人的经验和独立的思考，同时我当然也受到了很多其他人的观点的深刻影响，这一直可以追溯到我的小学时代，甚至更早。

关于心灵、大脑、模式、符号、自指和意识这些交织缠结的话题，很多知名的作者影响了我的思考。以下，大致按照他们影响我的时间先后顺序，我列出了对我影响最深的作者的感谢名单：欧内斯特·内格尔（Ernest Nagel）、詹姆斯·纽曼（James Newman）、库尔特·哥德尔（Kurt Gödel）、马丁·加德纳（Martin Gardner）、雷蒙德·斯穆里安（Raymond Smullyan）、菲佛（John Pfeiffer）、维尔德·彭菲尔德（Wilder Penfield）、帕特里克·苏佩斯（Patrick Suppes）、大卫·汉堡（David Hamburg）、阿尔伯特·哈斯托尔福（Albert Hastorf）、艾舍尔（M. C. Escher）、霍华德·德隆（Howard DeLong）、理查德·杰弗里（Richard C. Jeffrey）、雷·海曼（Ray Hyman）、霍妮（Karen Horney）、米克海尔·柏佳德（Mikhail Bongard）、图灵（Alan Turing）、格里格·蔡廷（Gregory Chaitin）、斯坦尼斯劳·乌拉姆（Stanislaw Ulam）、莱斯利·哈特（Leslie A. Hart）、罗杰·斯佩里（Roger Sperry）、雅克·莫诺德（Jacques Monod）、特里·维诺格拉德（Raj Reddy）、维克多·莱塞（Victor Lesser）、马尔文·明斯基（Marvin Minsky）、玛格丽特·波顿（Margaret Boden）、特里·维诺格拉特（Terry Winograd）、唐纳德·诺曼（Donald Norman）、艾略特·赫斯特（Eliot Hearst）、丹尼尔·丹内特（Daniel Dennett）、莱姆（Stanislaw Lem）、理查

德·道金斯（Richard Dawkins）、阿伦·威利斯（Allen Wheelis）、约翰·霍兰德（John Holland）、罗伯特·阿克谢尔罗德（Robert Axelrod）、吉尔斯·法库涅尔（Gilles Fauconnier）、保罗·波茨（Paolo Bozzi）、杰赛普·隆哥（Giuseppe Longo）、瓦伦蒂诺·布赖滕贝格（Valentino Braitenberg）、德里克·帕菲特（Derek Parfit）、丹尼尔·卡尼曼（Daniel Kahneman）、安妮·特丽斯曼（Anne Treisman）、马克·特纳（Mark Turner）和让·艾切森（Jean Aitchison）。这些作者的很多著作和文章都出现在参考文献中。多年以来，我有幸结识了其中数位，而我把与他们结下的友情视为生命中最大的欢愉之一。

就关系更加紧密的圈子而言，我与家庭成员、朋友、学生和同事之间的那些亲密交谈、电话、信件和电子邮件也在一生中不断影响着我。我也把这份感谢名单按照大致的时间顺序列出：南希·霍夫施塔特（Nancy Hofstadter）、罗伯特·霍夫施塔特（Robert Hofstadter）、劳拉·霍夫施塔特（Laura Hofstadter）、彼得·琼斯（Peter Jones）、罗伯特·柏宁格（Robert Boeninger）、查尔斯·布雷纳（Charles Brenner）、拉里·泰斯勒（Larry Tesler）、米歇尔·古德哈贝（Michael Goldhaber）、大卫·波利肯斯基（David Policansky）、彼得·史密斯（Peter Smith）、因加·卡尔林内（Inga Karliner）、弗朗西斯科·克拉罗（Francisco Claro）、彼得·林贝（Peter Rimbey）、保罗·柯松佳（Paul Csonka）、大卫·詹宁斯（David Jennings）、大卫·加斯特曼（David Justman）、斯科特·布莱希（Scott Buresh）、西德尼·阿尔克维奇（Sydney Arkowitz）、罗伯特·沃尔夫（Robert Wolf）、菲利普·泰勒（Philip Taylor）、斯科特·金（Scott Kim）、彭迪·卡纳瓦（Pentti Kanerva）、威廉姆·格斯皮尔（William Gosper）、唐纳德·拜尔德（Donald Byrd）、米歇尔·杜恩（Michael Dunn）、丹尼尔·弗里德曼（Daniel Friedman）、玛莎·梅雷迪斯（Marsha Meredith）、格雷·格罗斯曼（Gray Clossman）、安·特莱尔（Ann

Trail）、苏珊·万德尔（Susan Wunder）、莫大伟（David Moser）、卡罗尔·霍夫施塔特（Carol Hofstadter）、莱纳德·夏尔（Leonard Shar）、保罗·斯莫兰斯基（Paul Smolensky）、大卫·里克（David Leake）、彼得·苏贝尔（Peter Suber）、格雷格·胡波（Greg Huber）、伯纳德·格林伯格（Bernard Greenberg）、马雷克·卢果斯基（Marek Lugowski）、乔·贝克（Joe Becker）、梅拉尼·米切尔（Melanie Mitchell）、罗伯特·法兰西（Robert French）、戴维·罗杰斯（David Rogers）、本尼德托·西梅米（Benedetto Scimemi）、丹尼尔·德菲丝（Daniel Defays）、威廉·卡夫纳（William Cavnar）、米歇尔·加斯尔（Michael Gasser）、罗伯特·古德斯通（Robert Goldstone）、大卫·查尔默斯（David Chalmers）、加里·麦克格劳（Gary McGraw）、约翰·雷灵（John Rehling）、詹姆斯·马绍尔（James Marshall）、王培（Wang Pei）、阿吉尔·瓦尔茨（Achille Varzi）、奥利维尔罗·斯托克（Oliviero Stock）、哈利·方达里斯（Harry Foundalis）、哈米德·艾克比亚（Hamid Ekbia）、玛丽琳·斯通（Marilyn Stone）、凯莉·古特曼（Kellie Gutman）、詹姆斯·穆勒（James Muller）、亚历山大·林哈尔斯（Alexandre Linhares）、克里斯托弗·韦德曼（Christoph Weidemann）、纳撒尼尔·夏尔（Nathaniel Shar）、杰里米·夏尔（Jeremy Shar）、阿尔伯托·帕尔梅加尼（Alberto Parmeggiani）、阿莱克斯·帕西（Alex Passi）、弗朗西斯科·比安奇尼（Francesco Bianchini）、弗兰西斯科·拉拉达美尔（Francisco Lara-Dammer）、达米恩·苏利文（Damien Sullivan）、阿布西吉特·马哈巴尔（Abhijit Mahabal）、卡罗琳·斯特罗贝（Caroline Strobbe）、艾马努埃尔·桑德尔（Emmanuel Sander）、格兰·沃尔瑟（Glen Worthey）——当然还有我和卡罗尔的两个孩子，丹尼（Danny）和莫妮卡（Monica）。

印第安纳大学长期以来为我个人和我的研究小组（流动类比研究小组，简称“FARG”）提供了无比慷慨的支持，我对此深表谢意。大学里

几位重要的同人在过去20年的时间里维护着FARG的运转，他们是海尔加·凯勒（Helga Keller）、莫尔迪莫·罗文格鲁布（Mortimer Lowengrub）、托马斯·艾尔利希（Thomas Ehrlich）、肯尼斯·格罗斯·路易（Kenneth Gros Louis）、肯博·苏巴斯瓦米（Kumble Subbaswamy）、罗伯特·古德斯通（Robert Goldstone）、理查德·席福林（Richard Shiffrin）、米歇尔·杜恩（Michael Dunn）和安德鲁·汉森（Andrew Hanson）。他们所有人都是聪慧过人的伙伴和坚定不移的支持者，有的已经与我合作了数十年。我很幸运能与他们共事。

很长时间以来，我已经把我自己当作基本图书公司（Basic Books）的一分子了。我很感激那里的很多人在长达30年的时间里对我的照顾。最近几年，我一直在与威廉·弗卢赫特（William Frucht）紧密合作。对于他的思想开放性、他出色的建议和从不欠缺的热情，我真的深感钦佩。

有几个人为本书提供了巨大的帮助。肯·威利福德（Ken Williford）和尤赖亚·克里格尔（Uriah Kriegel）促成了这本书的启动；凯莉·古特曼（Kellie Gutman）、斯科特·布莱希（Scott Buresh）、比尔·弗卢赫特（Bill Frucht）、莫大伟和劳拉·霍夫施塔特提前阅读了部分草稿并给出了极为关键的建议；而海尔加·凯勒则为拿到各种许可而四处奔波。我感谢他们付出的一切。他们的付出远远超过了他们的本职。

上面提到的很多朋友，以及未及在此提及的朋友，形成了一片“云”，而我飘浮其上；有时候我把他们想象成一片“大都市区域”，而我正处于城市之内。每个人都有朋友。在这个意义上，我跟任何人都没有什么不同，但是这片云是我的云，它以某种方式定义了我。我为此感到骄傲，我为他们所有人而骄傲。所以我要用我的全部真心对这片朋友的云说：“感谢你们，感谢每一个人！”

序言

一位作家和他的书

直面意识的物理性

我从很小的年纪开始，就在思索自己的思维究竟为何物，以及推而广之，一切思维均为何物。我还记得自己曾努力地想要理解，我是如何编造出了那些一语双关的俏皮话、如何生发了那些数学观念、如何在讲话时出现了诸般口误以及如何凭空想象出了那些奇妙的类比，不一而足。我想知道做一名女孩子、身为说另一种语言的母语者、成为爱因斯坦、变成一条狗、一头鹰，甚至一只蚊子，到底是什么样子。大体而言，这段时期是一种令人愉快的存在。

在我12岁那年，一团深深的阴影笼罩了我的家庭。残酷的现实摆在我的父母、我17岁的姐姐劳拉以及我自己的面前：那时年仅3岁的茉莉，我们家最小的孩子，出了大毛病。没人知道那是怎么一回事儿，但是茉莉就是没法理解语言，也不能开口说话（她到现在依然如此，而我们始终没能找到症结所在）。她游刃有余地行于世间，甚至还带着一身优雅和魅力，可是她从不说话。这真是太令人难过了。

在很多年里，我的父母遍寻各种可以想见的途径，包括进行某种脑部手术的可能，但他们这种求索治愈之方或至少是某种解释的努力却只渐渐陷入越来越深的绝望之中。与此同时，我自己也对茉莉所处的困境展开了痛苦的思考，并想象着人们敲开我小妹妹的脑袋，窥视填充其中的神秘物质（这条途径到底也没有采用）。这种可怕的想法刺激我阅读了几本讲述人类大脑的入门书[注]。这些阅读于我的一生产生了重大的影

响，因为它迫使我第一次认真地思考意识以及作为一种存在或拥有物的“我”的物理基础[注]，它让我深感迷失、眩晕与极其的怪诞恐怖。

正是在那段时期，也就是我高中毕业的前夕，我与来自伟大的奥地利逻辑学家库尔特·哥德尔的神秘的元数学启示不期而遇，而且我还学会了如何编程。我使用的是斯坦福大学当时拥有的唯一一台电脑，巴勒斯220（Burroughs 220）。它被安置在破旧的“栎木学堂”（Encina Hall）的地下室里，那里昏暗隐蔽，恰合人意。我很快就对这台“巨大的机器大脑”上了瘾：它的橘色的指示灯以一种富有魔力的奇怪方式闪烁不定，揭示着它的“思想”；而它在我的指令下发现了优美的抽象数学结构，还用我正在学习的各种外语创作出古怪无常而荒诞不经的篇章。我与此同时还渐渐陷入了对于数理逻辑（symbolic logic）的痴迷之中：那些晦涩难解的符号踩着奇妙的魔法步点翩翩起舞，反照出种种真实、虚假、猜测、可能性与反事实性；而且我确信，经由数理逻辑，可以瞥见人类思想隐匿泉眼的深处。关于符号与意义、模式与理念、机器与心智、神经冲动与凡世灵魂等等一切想法不断翻腾，结果就是，我青春期的大脑里一片混乱。

幻景

这些观念的迷云，不仅在理智上吸引着我，也同样在情绪上牢牢地抓住了我。大约在我十六七岁的某一天，我正屏气凝神地冥思这些观念，忽然如梦初醒，而且自此以后始终认为，我们口中所称的“意识”，不过是一种幻景。可以肯定的是，这必然是一种相当特殊的幻景，因为这种幻景可以察觉到自身，而且显然它并不相信自己察觉到的是一种幻景；但不论如何，它终究还是一种幻景。这种名为“意识”的狡猾现象仿佛是倒提着自己的鞋带把自己举了起来，它仿佛是无中生有地创造了自身，然后每当有人想更仔细地观察它时，它又自行瓦解，回

归虚无。

我太过痴迷于弄清楚活着、作为一个人以及拥有意识到底是怎么一回事儿了，甚而感觉自己迫切地需要把飘忽易逝的思想努力地捕捉在纸面之上，以免它们飞逝溜走，一去不还。于是，我坐了下来，写下了一段发生在两名假想的同代哲学家之间的对话。我任意地将这两位哲学家命名为“柏拉图”和“苏格拉底”（我对真正的柏拉图和苏格拉底一无所知）。那也许是我笔下所写的第一篇严肃文字；不管怎么说，我是为它感到骄傲的，并从未将其丢弃。虽然现在回头再看我杜撰的这两名希腊哲学家所写的对话，未免太过幼稚和不堪，更别提它有多么粗陋；可尽管如此，我仍然决定把它作为引言收入本书，因为它蕴含了书中的诸多观念，而且在我看来，它可以为整本书奠定一种令人愉悦而又充满挑衅意味的基调。

投向裂缝中的一声呐喊

大约过了10年以后，我开始写自己的第一本书，当时设想的题目是“哥德尔定理与人类大脑”，首要的目标是把人类自我的概念与意识的谜团同哥德尔的震撼发现联系起来。哥德尔在一座令人望而却步的城堡的正中央发现了一个宏伟的“自我指涉”（self-reference）结构，尽管大胆的建筑师早已把一切自我指涉无情地驱逐在外（用我后来的说法，这就是一个“怪圈”）。哥德尔从无意义符号的基底上奇迹般地造出了自我指涉，而自我和灵魂也从由无生命物质组成的基底上奇迹般地现身，二者之间如此惊人的相似性令我确信，我们之所以能感知到“我”，秘密正在于此，而我的《哥德尔、艾舍尔、巴赫：集异壁之大成》这本书也由此诞生（并改了一个更抓人眼球的书名）。

这本书于1979年面世后，大获成功。事实上，该书作者此后的人生轨迹，很大一部分都得益于这次极大的成功。可不管这本书有多么通行

畅销，仍有一个问题始终困扰着我：《集异璧》（我总是这么称呼它，人们通常也习惯了这种叫法）这本书想要传达的根本性信息似乎被广泛地忽视不见。人们出于各种不同的原因喜欢上了这本书，却极少甚或没有人是看中了其最为核心的存在理由（raison d’etre）！光阴流转，我在自己后来出版的其他书中，屡次转引或补充了那份核心的信息，可我早在《集异璧》中已经切实在尽力陈明之义，似乎仍然没有收获多少理解。

1999年是《集异璧》出版20周年，基本图书公司的伙计们请为一个新出的特别纪念版作一篇序言。我很喜欢这个主意，所以就答应他们写了一篇。在这篇序言里，我讲述了有关这本书及其变迁始末的各种逸事，还特别描述了我对于这本书接受情况所感受到的绝望和沮丧。我在文末留下了这样一句抱怨话：“我有时候会感觉，我朝着一个空荡荡的裂缝呐喊出了一句自己深为珍爱的消息，结果无人听闻。”

不想到了2003年的春天，我有一日忽然收到来自两名年轻哲学家的一封相当友好的电子邮件。肯·威利福德和尤赖亚·克里格尔正在编纂的一套被他们称为意识的“自我指涉主义理论（或学说）”的选集，他们邀请我为这本书执笔贡献一个章节。他们热切地鼓励我加入，甚至还引用了我自己在那篇序言中发出的哀叹反过来劝说我，提出这可以给我提供一次真正能够改变一些事情的机会。他们对于我想传达的核心信息表达出真挚的兴趣，这让我发自内心地感激，而他们的热情也令我感动。同时，我也认识到，为他们的编书计划出一份力，确实是我的一个绝佳机会，可以借此再次尝试申明我的观点，而且面向的又是刚好合适的行家听众——钻研思维的哲学家。所以，接受他们的邀请，对我而言并不是一个很难做出的决定。

从雄壮的多洛米蒂山脉到温婉的布卢明顿

我在安泰塞尔瓦·迪·梅佐的一家恬静而朴素的旅店房间里开始写作我的章节。这个美丽的阿尔卑斯山村坐落于意大利的多洛米蒂山群，距离奥地利的边界线只有一箭之地。美好怡人的环境激发了我的灵感，我迅速挥笔写就了10页到15页之多，并以为自己大概已然完成了一半。然后，我返回印第安纳州布卢明顿的家中，继续坚持不懈地埋头写作。

完成这个章节的写作，花费了远比我预料之中更长的时间（我的一些读者应该能认出，这正是侯世达定律[注]（Hofstadter’s Law）的一个典型范例，这个定律说的是："做事所花费的时间总是比你预期的要长，即使你的预期中考虑了侯世达定律"）。更糟的是，这个章节的长度最终膨胀到了约定篇幅限制的5倍——简直是一场灾难！但是当肯和尤赖亚最终收到这篇文章时，他们对于我写的内容十分满意，对于我的轻率不慎也给予了最大限度的容忍；他们实在太过热切地想要让我加盟他们的编著了，于是表示他们可以接受加长篇幅的一章。特别是肯，他甘心付出额外的功夫，帮我把文章砍掉了一半。

与此同时，我开始意识到，我正在创作的内容绝不止于一本书的章节——它自己也可以成为一本独立的著作。于是，原本是由单一项目开始的工作，渐渐一分为二。我为自己写成的那一章文字所起的题目是"作为一个怪圈是什么样的"，呼应了心灵哲学家托马斯·内格尔（Thomas Nagel）讨论意识之谜的著名文章"作为一只蝙蝠是什么样的"[注]（What Is It Like to Be a Bat）；而我为自己准备创作的那本书起了一个更加简短也更为亲切的题目："我是个怪圈"。

肯·威利福德和尤赖亚·克里格尔合编的《意识的自我表征理论方法集》（Self-Representational Approaches to Consciousness）于2006年春面世。我的文章排在全书的最末，收在一个包含两章的板块之下，板块的标题是"超越哲学之外"（至于为什么说它超越哲学之外，实在也超越了我的认知，但不管怎么说，我是相当喜欢这个创意的）。我无法确定，在那种卓越而又相当专业化的设定下，这套观念能否对任

何人产生较大的影响；可我确实对自己的这本书抱以厚望，希望本书作为这些观念丰满完整且更加清晰显见的具象化形式，能够让这些观念触及形形色色的人，不论其居于哲学内外，不论老少，不论专家或门外汉，均可为他们提供一种关于自我和灵魂（当然包括那些环路模式！）的全新想象。无管怎么说，我都欠了肯和尤赖亚一个很大的人情。他们助我擦亮了本书创作最初的火花，而且一路上都给予我极大的鼓励。

就这样，在大约过了45年之后，（我的天哪！）我绕了整整一圈，再次拿笔书写起有关灵魂、自我和意识的文字，并迎面撞上了我在少年时代就曾首次体验过的那种未曾改变的神秘与恐怖。我还记得，那种令我们成为我们的物理法则，既糟糕透顶，又美妙惊人，10多岁的我在为此感到恐惧的同时，也痴痴地沉醉其中。

一位作家和他的读者

不要被这本书的标题迷惑，它讲述的并不是我，而是“我”这个概念。因此，它与你们、与读者之间的关系，与它和我之间的关系没有丝毫差异。我完全可以把题目改成“你是一个怪圈”。实话实说，要想更清晰地表达出这本书的课题和目标，我可能应该将其命名为“‘我’是一个怪圈”——可你还能想象出比这更厚实的书名吗？不如干脆把它改成“我是一个铅球”得了。

不管怎样，这本书讨论的是一个关于“我”为何物的庄严话题。那么它面向的读者应该是哪些人呢？同以往一样，我的写作总是为了影响那些接受过教育的普通公众。我几乎从来不为专家们写作，而且从某种意义上说，这是因为我自己也并不是一位真正的专家。噢，我要收回这句话；这样说是不公正的。毕竟，站在我人生的现阶段回首而视，我已经花了近30年的时间[注]，同我的学生们一起钻研类比推理与创造力的计算模型，观察并记录了各式各样的认知错误，收集了归类与类比的各种

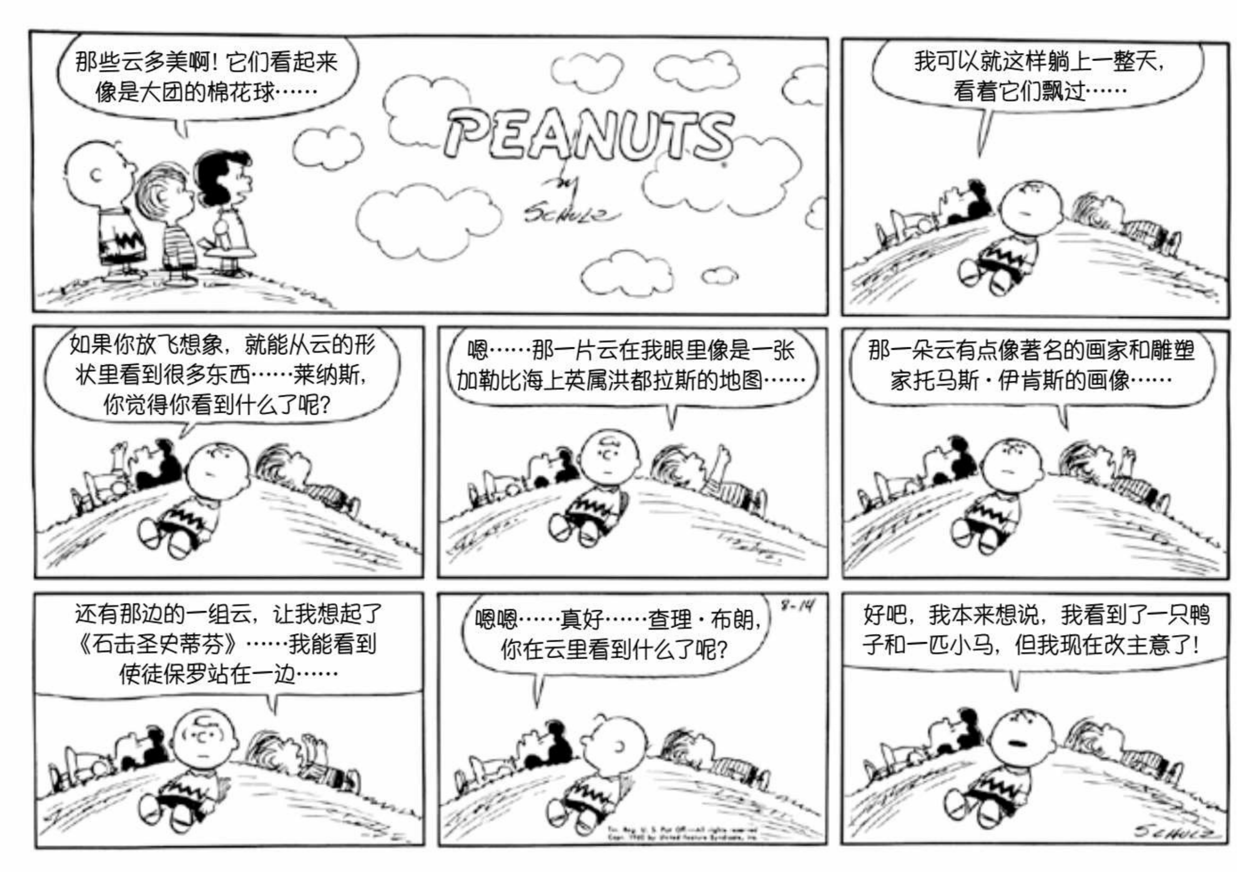
范例，研究了类比在物理与数学中的核心地位，深析了幽默的机制，推敲了概念创造与记忆提取的方式，并探索了词汇、习语、语言和翻译等所有的面向——而且在这30年中，我一直在讲授涉及思维的诸多方面以及关于我们如何认识世界的研讨课程。

所以，没错，说到底，我算是某种专家，而我的专长正在于对思考本身的思考。确实，正如我在前文所述，从我还是个10多岁的少年开始，我对于这个话题始终燃烧着热情。而我最为确信的结论之一便是：我们总是通过寻找并勾连与我们过去所知事物之间的类似性来进行思考的。因此，当我们利用大量丰富的例子、类比与隐喻之时，当我们回避了抽象的概括之时，当我们运用实际的、具体实在而简单朴素的语言之时，以及当我们直接谈论我们自己的经历之时，我们才能展开最有效的交流。

小马小狗教

多年以来，我都热衷于被我自己称为“小马小狗”的自我表达风格。这个说法的灵感来自著名卡通《花生》（Peanuts）中讨人喜爱的一话。我在这里节选了这集漫画。

我常常产生和最后一格中的查理·布朗（Charlie Brown）完全相同的感觉，就好像自己是一个想法与“在云间”扯不上半点关系的人，为人很实际，以至于为此感到难堪。我察觉到，有些读者对我形成了一种十分错误的印象，认为我具有一种极其渴望并且不懈追求最高级别抽象化的思维。我与此正好相反，而且我希望，阅读本书能让这一点更加明白无误。



至于我为什么把查理·布朗在此嘟囔的那句酸酸的话错记成了“小马和小狗”，我自己是浑然不知的，但这发生在很久以前的微改编，已经成为我个人言语中一个固定表达，所以不管是好是坏，它都成为我一直以来用以描述自己教学风格、演讲风格和写作风格的标准评语。

也许部分得益于《集异壁》这本书的成功，我在自己供职的两所大学里都幸运地享有了很大的自由度。印第安纳大学（大约25年时间）和密歇根大学（80年代的4年间）的慷慨包容令我得以奢侈地探索自己千变万化的兴趣课题，而不必承受“不发表则灭亡”的学术压力，或者还有更糟的追逐基金的无情重压。

我没有走上那种在专业杂志上一篇接一篇地发表论文的标准化学术路线。当然，我也发表过一些“真正的”文章，但是我最主要的精力都集中在以著书来表达自己的观点，而且这些书的写作向来着眼于清晰度

的最大化。

清晰、简明和具体，联合成为某种我所信奉的宗旨，包含了一整套需要时刻铭记的指导原则。幸运的是，很多善于思考的人都欣赏类比、隐喻和举例，赞同对行业术语的相对削减，而最后同样重要的是，他们认可从第一人称立场发出的言论。不管怎么说，这都是为了那些欣赏本书写作方式的人，而我所有的其他著作也都是如此写成的。我相信，这群读者不仅覆盖了业外人士与业余爱好者，也包括了众多专业的心灵哲学家。

如果你觉得，我在本书中讲了太多第一人称的故事，那并不是因为我沉迷于个人的生活之中，并自欺欺人地认为它相当重要，而只是因为这才是我最了解的生活，而且它还提供了形形色色的范例，我猜这足以反映绝大多数人生活的典型特征。我相信，大多数人都要通过听故事的方式才能对抽象的观念形成最清晰的理解，所以我在传达艰涩和抽象的观念时，总会尽量以我自己的生活作为媒介。但愿有更多的思想者使用第一人称的风格写作。

虽然我也希望以本书中的观点与哲学家们展开对话，但是我觉得自己的写作并不太像一名哲学家。在我看来，很多哲学家都以为，他们可以像数学家一样，切实地证明他们相信的观点，而为了达到这个目标，他们常常使用高度严谨和学术化的语言，有时候还会试图预测并反驳所有可能的对立论点。我对于这种自信深表钦佩，但我本人却不那么乐观，而且有点儿更偏向于宿命论。我认为一个人在哲学上不可能真正地证明任何东西；我认为一个人仅仅能说服那些本来就与其所宣扬的立场相当靠近的人们，而且很有可能这个人最终会去说服那些人。出于这种温和的宿命论，我在表达自己观点时所持的策略，更多地是以隐喻和类比为基础，而不去追求密不透风的严密性。事实上，这本书就是一碗盛满了隐喻和类比的沙拉。有些人会尽兴地享用我的隐喻沙拉，还有另外一些人会觉得它太……好吧，太过隐喻化了。但是我特别希望你，我亲

爱的食客，会发现这份沙拉正合你的口味。

最后的几句胡言乱语

我对待类比的态度是非常严肃的，甚至于不惜千般麻烦，为我这份“沙拉”中大量的类比一一做了索引。因此，在索引中列举范例时，存在两类主要的标题。

一类是“类比，严肃的例子”；另一类是“一次性类比，随意的例子”。我之所以做出这种古怪的区分，是因为，虽然我的很多类比对于观点的传达都是至关重要的，但也有一些只是为了添加点儿佐料。这还说明了另一个道理：归根结底，这本书（或任意一本书）中的每一个想法，几乎都是一则类比[注]，因为它必然涉及将某种事物识别为另一种事物的变形。所以，每次当我写下“类似地”或“相比之下”的时候，都意味着一次隐含的类比，而每当我选定一个单词或短语（如“沙拉”“仓库”“底线”）时，我都是在与储存在自己生活经验仓库中的某物进行类比。底线是，此中的每一个想法都可以列于“类比”之下。尽管如此，我还是忍住没把自己的索引如此细化。

我对于这本书最初的设想，是对《集异璧》的核心信息进行一次过滤性重述，筛掉几乎所有的形式化标记，也不再沉迷于普希金式的离题[注]而游走在禅宗佛教、分子生物学、递归运算、人工智能等等千差万别的话题之间。换言之，我以为自己已经在《集异璧》和我的其他著作中充分地阐述过我想要在此书中（重新）阐述之义，但令我倍感惊喜的是，在我开始动笔之后，新的观点如雨后春笋般冒出地面，四处萌生。这让我深感快慰，也使我感觉到自己的新书不仅仅是一本（或几本）早年之作的旧调重弹。

《集异璧》成功的关键之一，是它在章节与对话之间的穿插交替，

但是30年后，我不想再抄袭我自己的这种穿插布局了。我如今已经换了一种不同的思维框架，而我希望这能够在本书中反映出来。但是当我的写作越来越接近尾声时，我开始想要把我的观点和心灵哲学中的常见观点进行比较，所以我开始写出类似这样的话："怀疑者也许会做出如下回应……"在我已经数次写下了这种表述之后，我意识到自己在无意中又开始编写起发生在我自己与一名假想的持怀疑立场的读者之间的对话，于是我捏造出两个名字奇怪的人物，让他们展开对攻，结果这成为本书中篇幅最长的章节之一。这段对话本无意于制造令人捧腹的滑稽效果，但我也希望我的读者能在阅读中时不时地偶尔一笑。不管怎么说，热爱对话形式的读者，请打起精神来——这本书里有两段对话。

形式与内容的交互是我的一生所爱，这本书自然也不例外。就像我之前出版的几本书一样，我有机会以最精密的细节程度完成它的排版(注)，而我在每一页上对于视觉优雅性的追求，也曾无数次地让我调整了表述观点的措辞方式。这在有些人看来或许有点儿本末倒置，可我认为，任何人的写作都能通过对形式的在意而获得提升。我希望，阅读本书不仅是一次智力上的挑战，还是一场美妙而愉悦的视觉体验。

一种有用的青春

《集异璧》的作者是一个相当年轻的人（开始动笔时，我只有27岁；而到了28岁，我已完成了第一稿——全部都是用钢笔在画线稿纸上手写的）。虽然年纪轻轻的我已然经历过属于我自己的那一份应得或不应得的苦难、悲伤与道德上的灵魂求索，可是在那本书中，生活的这些面向却没有得到太多的提及。而写作本书之人所亲尝过的苦难、悲伤与灵魂求索，明显相较甚多，笔端也更加频繁地触及生活的这些面向。我想这正是变老的一种表现——一个人的写作变得更向内转，更爱反思，也许更具哲理，却也可能只是徒增悲伤。

很久以来，我一直深深迷恋着小说家安德烈·马尔罗（André Malraux）的著名小说《人的境遇》（La Condition humaine）的诗意标题。至于这个引人遐思的短语究竟意味着什么，我猜每个人都有自己的想法。我愿意以《我是个怪圈》作为我自己描述“人的境遇”含义的全力一试。

在我最喜欢的《集异壁》封面简介之中，有一则是出于物理学家兼作家杰里米·伯恩斯坦（Jeremy Bernstein）的惠赠，其中一句评语是：“它拥有一种青春的活力和一种美妙的才华……”这在我听来，真是声声入耳，如遇知音！然而不幸的是，这句溢美之词却不知何时遭到篡改，结果就是，如今在市面上横飞的几千本《集异壁》，封底都印着伯恩斯坦的这样一句赞誉：“它拥有一种有用的活力……”相比于一种“青春的”活力，这是多么令人泄气啊！可是话说回来，我这本表现得更加老成也更为清醒的新书，很可能也会在以后的某一天在某处被某人形容为具有一种“有用的”活力。我猜这也算不上是对一本书最差的评论。

下面，我要停止继续谈论我的书了；我想让我的书自己说话。我希望你能在这本书中发掘出有趣而新奇的信息，甚至还带有一种哪怕不再青春却依然有用的活力。我希望，阅读这本书能让你以一种全新的方式反思作为一个人的存在意味着什么——事实上，就是反思纯粹的“存在”之义。而且我还希望，当放下这本书时，你也许也会把自己想象成一个怪圈。果真如此，我将不胜欣慰。

——印第安纳，布卢明顿

2006年12月

---

1. 主要是菲佛（Pfeiffer）、彭菲尔德（Penfield）、罗伯茨（Roberts）以及伍尔德里奇（Wooldridge）的作品。

2. Placing commas and periods outside quotation marks when they are not part of what is

being quoted exhibits greater logic than does American usage, which puts them inside regardless of circumstance. In this book, the logical convention (also the standard in British English) is adopted.

3. 出自《集异璧》（1979）。

4. See Chapter 24 in [Hofstadter and Dennett].

5. See, for instance, [Hofstadter and Moser], [Hofstadter and FARG], [Hofstadter 1997], and [Hofstadter 2001].

6. See [Hofstadter 2001].

7. See James Falen's sparkling anglicization of Pushkin's classic novel-in-verse *Eugene Onegin*[Pushkin 1995], or see my own translation [Pushkin 1999], There is no sublimer marriage of form to content than *Eugene Onegin*.

8. 在这本书的排版中，我的主要关注点在于分段的处理，其中最基本的一个原则是：如果一个自然段在分页的时候被分开，那么要确保该自然段在分开的两页中都不出现“单行”的情形。另一条指导性原则是，我希望排版出来的文稿看起来很舒适，不要太挤，但也不要太松（在电脑排版的今天，这种情形经常发生，很是令人头疼）。为了避免这些问题，我经常会对文稿进行文字上的增删，来让读者获得阅读的舒适感。

引言

一场友善的争论

正如我在序言中所叙，写下这段对话时，我还是个10多岁的少年。这是年轻气盛的我与这些复杂的观念展开缠斗的第一次尝试。

剧中出场人物：

柏拉图：一名真理的追求者，怀疑意识是一种幻觉。

苏格拉底：一名真理的追求者，相信意识的现实性。

* * *

柏拉图 可是，苏格拉底啊，你所谓的“生命”又意味着什么呢？在我看来，一个生命就是一具身体，它在诞生之后，成长、吃喝、学习如何应对各种不同的刺激，最终获得生殖繁衍的能力。

苏格拉底 柏拉图啊，你认为一个生命就是一具身体，而并非拥有一具身体，我觉得这有趣极了。因为可以肯定的是，如今有很多人都会说，至少某些生命，在身体之外还拥有独立的灵魂。

柏拉图 没错，而我也赞同他们的说法。我本就应该说，生命拥有身体。

苏格拉底 那么你也就同意，不管多么微不足道，跳蚤和老鼠也是拥有灵魂的喽？

柏拉图 我的定义确实能推出这一点，没错。

苏格拉底 那么大树也有灵魂吗？花花草草呢？

柏拉图 你这是在用文字游戏为难我，苏格拉底。我要修正我的说法——只有动物拥有灵魂。

苏格拉底 非也非也，我可不是仅仅在跟你玩文字游戏。如果你去观察足够小的生物，就会发现，动物与植物之间并没有什么分别。

柏拉图 你的意思是，有某些生物，它们同时享有植物与动物的属性？没错，我猜我是可以想象出这样一种东西的，就是我自己嘛。我想，现在你要强迫我说出，只有人类拥有灵魂。

苏格拉底 不，恰恰相反，我要问你，你通常认为哪些生物拥有灵魂呢？

柏拉图 咳，所有的高等动物呗——那些有能力思考的生物。

苏格拉底 好吧，高等动物至少都是活着的。那么，你能真心地认为一株草是像你自己一样的生命吗？

柏拉图 苏格拉底啊，我这么跟你说吧：在我的想象中，只有拥有灵魂的才是真正的生命，所以我必须把小草摒弃在真正的生命之外；但是我可以说，它具有生命的特征。

苏格拉底 我明白了。所以你会把没有灵魂的生物归类为只是看上去像是活着之物，而拥有灵魂的生物则是真正的生命。那我可不可以说，你那个“何为真正的生命？”的问题，取决于对灵魂的理解？

柏拉图 对，没错。

苏格拉底 而且你也说过，你认为灵魂就是思考的能力？

柏拉图 是的。

苏格拉底 那么你真正在寻找的，其实是“何为思考？”这一问题的答案。

柏拉图　苏格拉底啊，我跟上了你论证的每一步，可得出的结论却让我很不舒服。

苏格拉底　这可不是我的论证啊，柏拉图。你提供了所有的事实，而我不过是从这些事实中逻辑地推出结论而已。当自己的观点由他人之口陈述出来的时候，人们常常就不再信任它了，这可真是奇妙。

柏拉图　你说得没错，苏格拉底。而且，要解释清楚何为思考，肯定不是一个简单的工作。在我看来，最纯粹的思想似乎就是对某个事物的知识；显而易见，知道某物，远不止于把它写下来或者说出来。如果一个人知道某件事物，那么这些动作都可以完成；而且，一个人还可以通过听到关于它的言说或者看到关于它的写作而知道这件事物。可是，知道远远不止于此——它还是信念——我在这里只使用了它的一个同义词。苏格拉底啊，我发现，到底什么是知道，已经超越了我的理解力之所及。

苏格拉底　柏拉图啊，这个想法真是有趣。你是不是在说，知道并不像我们以为的那么熟悉？

柏拉图　是啊。因为我们人类拥有知识，或者说信念，所以我们才得以成为人。可是当我们试图对知道本身进行分析时，它却遁走无形，离我们远去了。

苏格拉底 这么说，我们是不是最好对我们所谓的“知道”或“信念”多留个心眼，别再将它们视为理所当然之物了？

柏拉图 正是。我们在说出“我知道”时要倍加小心；而当我们的头脑想让我们说出“我知道”的时候，我们也必须认真思考，这么说究竟意味着什么。

苏格拉底 的确如此。如果我问你：“你活着吗？”你会毫无疑虑地回答说：“是的，我活着。”而如果我问你：“你怎么知道你活着呢？”你会说：“我感觉得到，我知道我活着——确实，难道活着的状态不就是知道并感觉到一个人正在活着吗？”我这么说对吗？

柏拉图 非常对。我也说不出比这效果更好的话了。

苏格拉底 现在，我们假设有一台能够造句并回答问题的机器被制造出来。假设我问这台机器：“你活着吗？”并假设它给我的回答跟刚才你给我的答案一模一样。那么你认为它的这个回答，有效性如何？

柏拉图 首先，我会提出抗议，指出任何机器都不可能知道什么是词语，也不可能知道词语的意义。[注] 一台机器只能用一种机械方式来处理词语，这跟装罐机器把水果装到罐头里面没什么不同。

苏格拉底 我会驳回你的抗议，理由有二。第一，你肯定不会主张，人类思想的基本单元就是词语吧？因为众所周知，人类具有神经细胞，而神经元的运转法则是算术式的。[注] 第二，你在此之前刚刚提醒过我们，要谨慎使用“知道”这个动词，可你自己却在此相当肆无忌惮地使用了它。你凭什么说，任何机器都永远不可能“知道”什么是词语以及词语的意义呢？

柏拉图 苏格拉底，你的意思是，机器也能像我们人类一样知道事实吗？

苏格拉底 你刚刚声明过，你自己甚至都无法解释知道到底是什么。你在小时候是怎么学会“知道”这个动词的？

柏拉图 显而易见，我是在身边使用它的人之中耳濡目染地就学会了。

苏格拉底 也就是说，你掌握它是通过自动化的行为。

柏拉图 不是……好吧，也许我知道你是什么意思。我渐渐习惯了在特定的语境下听到这个词，于是自己也能在那些语境中使用它，这或多或少是一种自动化的方式。

苏格拉底 就跟你现在使用语言差不多——不需要对每个单词反复思量？

柏拉图 没错，正是如此。

苏格拉底 现在，如果你说：“我知道我活着。”这句话仅仅是从你大脑中输出的一个反射，而不是有意识的想法的产物。

柏拉图 不，不对！我们两个人中，必有一人犯了逻辑错误，不是你，就是我。并不是我说出口的所有想法，都仅仅是反射行为的产物。有些想法，我在说出口之前是有意识地思考过的。

苏格拉底 你说你有意识地思考它们，是什么意思？

柏拉图 我不知道。我想是指我努力找到描述它们的正确词语。

苏格拉底 是什么在引导你找到正确的词语呢？

柏拉图 嗯，我会遵循逻辑去搜索我熟悉的同义词、相似词等等。

苏格拉底 换句话说，是习惯引导了你的想法。

柏拉图 对。我的想法是由把词与词系统化连接在一起的习惯所引导的。

苏格拉底　那么还是一样，这些有意识的想法依然是反射行为的产物。

柏拉图　虽然我跟上了你的论证，但若是这样的话，我怎么也想不明白，我如何才能知道自己是有意识的，又如何能感觉到自己正在活着。

苏格拉底　可是这番论证本身就表明，你的反应仅仅是一种习惯，或者说反射行为，而且没有任何有意识的想法在指引你说出，你知道自己活着。你可以停下来好好想想，你真的理解自己说出这样一句话是要表达什么意思吗？或者说，这句话是不是没有经过你任何有意识的思考，就直接出现在你的脑子里了？

柏拉图 真的，我实在有点晕头转向，我几乎什么也不知道了。

苏格拉底　实在有趣，我们看到了一个人的头脑在沿着另一个轨道运转时是如何失灵的。你知道你对那句“我活着”的理解有多么贫乏了吗？

柏拉图　　是的，我必须承认，这真的是一个没那么容易理解的句子。

苏格拉底　我觉得我们很多行为的发生，都跟你说出那个句子的情况是一样的——我们以为它们是从有意识的想法中生发而来，但若仔细加以分析，那个想法的每一分每一寸都是自动化的，而且没有任何意识的参与。

柏拉图　这么说，感觉到一个人活着，其实是一个错觉，它来自催促一个人不做任何理解而开口说出这句话的反射。如此而言，一个真正的生命便简化为一个复杂反射的集合。苏格拉底啊，这样一来，你已经告诉了我生命到底是什么。



1. 持这一个观点的当代代表人物是约翰·塞尔。

2. 这也同时意味着，只要机器能够模仿神经元的算术运转方式，其行为特征便可以与人类或者动物毫无差别。换句话说，人工智能在这个意义上是可能的。然而即便如此，这种人工智能和以通过字词和概念在特定规则下对思想的流动——而不是物质性的硬件——进行表征的系统存在很大的差异。

第1章

论灵魂及其尺寸

灵魂的碎片

1991年年初，一个阴郁的日子。我站在父母家的厨房里，我的母亲正在看一张父亲的照片。照片甜蜜而动人，大约摄于15年前，而我的父亲数月前刚刚过世。母亲用一种绝望的口吻对我说：“这张照片还有什么意义呢？什么也没有。这就是一张四处撒着暗点的平板纸。一点儿用都没有。”母亲充满哀伤的话语中渗透的悲凉，让我的脑子打起转来，因为我本能地知道，我是不赞同她的，可是我还不太清楚要如何告诉她这张照片在我看来应该以何种方式加以看待。

我经历了几分钟情绪激烈的沉思，或者毫不夸张地说，是灵魂的拷问。然后，我忽然想到了一个类比，我感觉它足以向我的母亲传达出我的观点，而且也希望它能带给她哪怕只有一点点的宽慰。我当时对她所说的话与下面这一段大致相似。

“我们在客厅里有一本肖邦的钢琴练习曲。每一页都跟爸爸的这张照片一样，是二维的、平板的、可以折叠的纸张，纸上也只是撒着些暗色的记号——但是，想一想它们在过去的150多年里对全世界的人所带来的巨大影响吧。多亏了那些在平板纸面上散布的黑色记号，才会有不知数千个人，共同付出数百万个小时，循着复杂精妙的模式在钢琴键盘上舞动手指，创造出为他们带来难以描述的享受与伟大意义感的声音。这些钢琴家转而又向包括你和我在内的数百万名听者传达了那曾拨动弗里德里克·肖邦心弦的深刻情绪，让我们每个人都得以或多或少地触及

肖邦的内在——脑海中的生命经验，或者不如说是弗里德里克·肖邦的灵魂。而那些纸面上的记号，不啻灵魂的碎片——弗里德里克·肖邦的灵魂的零落残余。这些音符奇怪的几何形状，每一种拥有一股独特的力量，可以在我们的头脑中复活另一个人内在经验的细碎片段——他的痛苦、他的愉快、他最深的热情与不安，我们也借此至少部分地感受到了成为那个人是什么样的感觉，还有很多人就此炽烈地爱上了他。我们作为与爸爸亲密相知的人，看他的照片，也会以同样有效的方式，召回关于他的微笑与温柔的最为清晰的记忆，并在我们这些活人的头脑中，激活爸爸留存在我们脑海中的某些最为核心的表征。他的灵魂的小小碎片再次翩翩起舞，只是借用了他人的头脑作为场地。同肖邦练习曲的一篇乐谱无异，那张照片也是某个逝去之人的灵魂的碎片。而我们只要还活着，就应该对此倍加珍惜。”

虽然上面这段文字的辞藻要比我当时对母亲说的话更加讲究一点，但它讲出了我想要传达的信息的本质。我不知道母亲对于那张相片的感觉有没有因为那段话而发生什么变化(注)，我只知道那张照片依然还在那里，摆在她厨房里的一个柜台上。每一次当我看到它时，我都会想起那次对话。

作为一个西红柿是什么样的？

我把西红柿切开、吞下，没有丝毫的负罪感。我不会因为吃掉了一个新鲜的西红柿而寝食难安。我从没想到过要问问自己我吃的是哪一个西红柿，或者我是不是在吃掉它时掐灭了一缕内心之光，我也并不认为努力想象西红柿被切开并盛在我的盘子里时是什么感觉是有意义的。于我而言，西红柿就是一个没有欲望、没有灵魂、没有意识的实体(注)，我可以毫无顾虑地任意处置它的“身体”。确实，西红柿除了身体之外一无所有。对西红柿而言，没有“心身问题”可言。（亲爱的读者，我希望我

们在这一点上能达成一致！）

在我拍死一只蚊子的时候，也不会有任何迟疑，虽然我会尽量避免踩死蚂蚁，而且当房子里出现蚊子以外的昆虫时，我通常都会尽量把它捉住，转移到室外，把它完好无损地放生。我有时候会吃鱼和鸡[注]，但我在很多年前就已经不吃红肉了。不要牛肉、不要火腿、不要培根、不要午餐肉、不要猪肉、不要羊肉——不用了，谢谢你，女士。需要注意的是，我依然喜欢一份BLT（火腿、莴苣、番茄三明治）或者一个美味汉堡的味道，只不过出于道德的原因，我不再吃它们了。我并不想在这里没完没了地推销，但是我的确需要讲一点点我在素食主义中的所学，因为它们和灵魂的关系太大了。

豚鼠

我在15岁的时候找了一份暑期工作，是在斯坦福大学生理实验室的那台弗里登（Friden）机械计算器上敲打按键。（在那个年代，整个斯坦福校园里只有一台电子计算机，而且绝大多数的科学家甚至都不知道到它的存在，更别说想起来用它帮自己计算了。）连续数小时的“敲数字”是一项非常累人的工作，而我做的所有工作都是为了一个名叫南茜的研究生的研究项目。有一天，她出于体恤之心，问我愿不愿意尝试一下实验室里其他类型的工作。我回答说：“当然了！”于是，那天下午，她护送我走上了生理学大楼的第四层，向我展示了他们关养用于实验的动物的笼子——准确地说，那些动物就是豚鼠。我至今犹记得那股刺鼻难闻的味道，还有那些橘色皮毛啮齿类小动物急匆匆地窜来窜去的景象。

第二天下午，南茜客气礼貌地问我愿不愿意帮她从顶楼取两只她下一轮实验要用的动物下来。然而，我根本没有机会给她一个回应，因为我刚刚开始想象自己伸手到那些笼子里去选择两只将被扼杀的毛绒小生

命，就立刻头晕目眩，天旋地转，顿时晕倒在地，脑袋撞到了混凝土地面上。我记得的下一件事，就是自己向上看到了实验室主管乔治·费根（George Feigen）那张写满深切担忧的脸。他是与我们家关系亲密的老朋友，担心我在摔倒时伤到了自己。幸运的是，我毫发无损。我慢慢地站起身，然后骑着我的自行车回家，那天下午就这样过去了。自此以后，再也没有人要求我为了科学的缘故而去挑选以供牺牲的动物。

猪

足够奇怪的是，尽管与夺走一个生灵生命的想法发生了极其令人不安的正面相遇，我在随后的几年里还在继续吃汉堡包和其他种类的肉食。我觉得自己没有太多思考过吃肉这件事，因为没有哪位朋友思考过这件事，也自然没有人谈论过它。对于我认识的每一个人而言，吃肉只不过是生活中的一个背景事实。另外，我不无羞愧地承认，在那些日子里，"素食主义者"这个词在我的心中唤起的联想是一个严词说教的疯子意象，电影《七年之痒》（*The Seven Year Itch*）中有一个发生在曼哈顿素食餐馆的场景，将这一刻板印象表现得淋漓尽致。可是到了我21岁的某一天，我读到了挪威籍英语作家罗尔德·达尔（Roald Dahl）的短篇小说《猪》[注]，而这篇小说对我的人生产生了深远的影响——而且经由我，也影响了很多其他生物的生命。

《猪》有一个轻松而诙谐的开篇。一位名叫列克星敦的天真的年轻人，从小被他的葛洛斯潘阿姨（Aunt Glosspan）[注]培养成了一位严格的素食主义者。在阿姨死后，他发现自己爱上了肉的味道（虽然他不知道自己吃的东西是什么）。很快，就像在达尔所有的小说里一样，事情开始诡谲地扭曲起来了。

列克星敦不知道这种名为"猪肉"的美味是何种食材，于是在好奇心的驱使和一位新朋友的引荐下，决定前往一家屠宰场游览参观。我们和

他肩并肩地坐在等候室里，还有其他的参观者也在。他漫不经心地看着等候室里各种各样的人被一个接一个地叫到名字，开始参观。最后终于轮到了他。他被领出了等候室，进入了脚镣区。在那里，他看到猪被倒提着悬吊起来，后脚穿在钩子里，挂在一条正在移动的链条上。它们的喉咙被割开，血液喷射出来，然后继续猪头朝下，进入“拆解线”，落到盛有滚烫开水的大锅里，烫除体毛之后，它们的头和四肢被肢解拆分，也为开膛破肚做好了准备。然后，它们被分装在整洁的小包装纸袋里，发送到全国各地的超市，同其他的玫瑰色竞争者同坐在玻璃柜里，等待购买者的称赞，并希望他们能把自己买回家中。

正当列克星敦以一副事不关己的着迷观察这一切时，他自己的脚忽然被猛拽起来，大头朝下颠倒过来。他发现自己跟他一直在观看的那些猪一样，也被悬吊在那条移动的链条上了。他再也没法保持平静了，大叫起来：“这个错误太可怕了！”可工人们充耳不闻。很快，链条就把他拉到了一个面容和善的小伙子跟前。列克星敦希望他能意识到眼前状况的荒诞性，可事与愿违，这位温文尔雅的“行刺者”（sticker）抓住了列克星敦的耳朵，把倒挂的年轻人拉近了一点，然后带着慈爱的微笑，用一把锋利的小刀的刀刃，娴熟地割开了少年喉部的血管。年轻的列克星敦继续这场令人匪夷所思的颠倒旅程。他强壮有力的心脏把血液泵出他的喉咙，洒在混凝土地面上，而即便他大头朝下，意识正在迅速地流失，他还是隐隐约约地觉察到，在他前面的那些猪，正一头接一头地掉入蒸汽氤氲的大锅之中。说也蹊跷，它们之中有一头似乎在两只前蹄上戴着白色手套的猪，这让他想起了刚才把他从等候室领入参观区域的那位戴着白手套的年轻女人。揣着最后一丝迷惑不解的念头，列克星敦迷迷糊糊地从这个“所有可能的世界中最好的一个世界”滑入了下一个世界。

《猪》的收尾场景在我的头脑里久久回荡。我的脑海中不断闪现出两种意象，来来回回，摇摆不定，一会儿把自己想象成一头倒挂在钩子上哼哼呼噜的猪，一会儿又想象自己是滑入大锅的列克星敦……

恶心，明心，革心

这则故事的阴霾直到一两个月后依然挥之不去。那时，我的父亲为了参加一个物理学会议，已经带着我、母亲和姐姐劳拉来到了山路崎岖的撒丁岛（Sardina）南端的卡利亚里城（Cagliari）。会议组织方想要以一种奢华的本土方式谢幕，于是在卡利亚里城外的一座公园里策划了一场盛大的宴会。宴会上，会当场烤一头乳猪，并当着所有用餐者的面切肢片肉。我们作为会议的嘉宾，都被邀请参与这一饱受尊崇的撒丁岛传统。可是我最近才读过达尔的小说，还深深地处于它的影响之下，甚至无法设想自己参与到这样一个仪式当中。在我的新的思维框架之中，我甚至都难以想象竟然有人会愿意出席，更别提分食乳猪的身体了。没想到的是，我的姐姐劳拉一想到那个即将上演的场面，也被吓坏了。于是，我们两个便留在了宾馆里，开心地吃了些意大利面团和蔬菜。

挪威人的《猪》和撒丁岛的乳猪，让我遭受了两连击，结果让我在我姐姐的带领下，完全放弃了肉食。我还拒绝购买皮鞋和皮带。很快，我就变成了一个狂热的说教者，四处宣扬自己新的教义。我曾经设法劝服几个朋友吃了几个月的素，我还记得自己当时是多么心满意足。可是让我失望的是，他们都逐渐地放弃了。

在那些日子里，我经常纳闷：我的某些偶像——例如，爱因斯坦——怎么也会是个肉食主义者呢？我找不到任何解释。我最近通过网络搜索，发现爱因斯坦事实上在感情上是倾向于素食主义的，而且这不是出于健康的考虑，而是出于对生灵的同情。虽然这让我大感宽慰，但我当时可是不了解这些事实的，而且不管怎么说，还有很多我崇拜的其他英雄是彻头彻尾的肉食主义者，而且他们完全清楚自己在做什么。这些事实令我悲伤，也让我陷入了迷惘。

再反转，再进化

特别奇怪的一件事是，只过了几年，我本人也因为觉得美国社会的日常生活压力实在太过强大，所以放弃了自己一度投入了全副热情的素食主义。而且在一段时间内，我所有那些强度极高的反思都彻底伏于地下。我想，20世纪60年代的那个我会觉得这次反转是完全不可理喻的，可是这两种版本的我确实都在同一个身体内存在过。我真的还是同一个人吗？

就这样又过了几年，仿佛我几乎从未经历过任何顿悟。但是后来有一天，当我还是印第安纳大学的一位初级助理教授时，我遇到了一名具有高度思考力的女人。她和我一样，也接受了同样的素食主义哲学，也出于类似的原因开始吃素，但是她比我坚持的时间更长。苏和我成了好朋友，我也很钦佩她所持立场的纯粹性。我们的友谊引发我又一次彻头彻尾地思考了这件事，而我旋即重返回到自己那个绝不杀生的“后《猪》”立场。

在接下来的几年里，又出现过几次更多的波动，但到了30多岁的时候，我终于抵达了一个稳定的状态，找到一种折中的方案来呈现我进化中的直觉——灵魂有不同的尺度。即便在我看来这丝毫也不明晰，但我还是愿意接受一种相对模糊的观点，即认为某些“足够小”的灵魂为了满足“更大的”灵魂的欲望而被牺牲，是合情合理的，比如为了我的灵魂或者其他人类的灵魂。虽然把分割线划在哺乳动物上明显有点失之武断（这类的分割线一定如此），可这却成了我新的信条，而我也在随后20多年里坚守未变。

无生命的肉的谜团

我们这些说英语的人不吃猪或牛（pig or cow）；我们吃的是猪肉和牛肉（pork and beef）。我们的确吃鸡肉（chicken），但我们不吃鸡（chickens）。有一次，我的一位朋友的年幼女儿欢快地向她的父亲发

出惊叹说，某种咯咯叫的会下蛋的家禽同她经常在晚餐的餐盘里看到的食材，是用同一个词来表示的。她觉得这是一种极为滑稽的巧合，就像“calf”这个词既表示牛犊又表示腿肚子这个巧合同样有趣。不消说，当她被告知，那种美味的食物与咯咯叫的下蛋动物其实是完全相同的一种东西时，她是相当沮丧的。

想来，我们在小时候大概都经历过同样的困惑，发现自己正在吃那些在我们的文化里据说是相当可爱的动物——绵羊、兔宝宝、牛犊、小鸡，诸如此类。我还朦朦胧胧地记得自己关于这一谜团的真实的童年困惑，可既然吃肉是如此稀松平常之事，我通常都会把这种困惑晾在一边，没有给予太多的思考。

尽管如此，食品杂货店却总以一种特别恼人的方式把这个议题鲜明生动地推到幕前。巨大的玻璃陈列柜里，展示着各种各样泛着不同奇怪颜色的看上去黏糊糊的肉团，标签上写着“肝”、“肚”、“心”和“肾”，有时候甚至是“舌”和“脑”。这些不仅听起来像动物的身体部位，而且*看起来*也很像。幸运的是，名为“碎牛肉”之物看上去并不那么像动物的一个部位，而我之所以说这是“幸运”，是因为它实在太好吃了。我可不想被劝说戒掉*这个*！培根也美味极了，而且一条一条的那么瘦，做熟了以后爽脆可口，让人几乎没法跟动物联想到一起。这是多么幸运啊！

解开这个谜团的地方正是食品杂货店后门的卸货口。有时候，一辆大卡车会停在那里，当卡车的后门弹开之后，我看到整块的肉和骨头毫无生气地垂挂在骇人的金属钩子上，悠来荡去。我抱着一种病态的好奇心目送这些尸骸被抬入店铺后面，挂在头顶轨道的挂钩上，好方便他们来回移动。所有这一切都让处于青春期前的我感到十分不安。每当我看到一具尸骸时，我都忍不住默想，“那个动物以前是谁呢？”我心里猜想的不是它的*名字*，因为我知道家畜们没有名字；我试图抓住的是某种更加哲学化的东西——作为*那只*动物与某只*其他*的动物有什么不一样的感觉。当这只动物遭到屠杀时，突然熄灭的那道独一无二的内心之光是

什么？

我在10多岁的时候曾前往欧洲，让这个问题更加鲜明地突显了出来。在那里，无生命的动物身体（通常是剥了皮的，无头无尾，但有时也不是）被展示在所有顾客的面前。我最鲜活的记忆是关于一家食品杂货店的，它在圣诞节假期前后，在走廊中央的桌子上摆了一个切下来的猪头。如果你恰好从后面接近它，你就会看到一个扁平的十字交叉切口，展示着那头猪脖子的所有内部结构，就和它刚刚被斩首时的样子一模一样。那里面密密麻麻地分布着交叉的线条，曾经正是这些线条把个体身上所有分散的部位连接到了它头部的中央“指挥部”。从另一侧看，则这头猪的脸上挂着一种仿佛是冻住的微笑，给我一种毛骨悚然的感觉。

我又一次禁不住地纳闷，“那个脑子里曾经住过谁？谁曾经活在那里？谁曾经透过那双眼睛向外看，通过那些耳朵听声音？这一大块肉曾经究竟是谁？它是雄性还是雌性？”当然，这一切都是没有答案的，而且别的客人似乎并没有注意到这处展示。在我看来，似乎除我以外不再有人直面关于生存、死亡和“猪的同一性”的尖锐问题，而这只沉默的、一动不动的猪头却如此强力而搅扰地把这个问题从我的心里激发了出来。

有时候，我在碾死了一只蚂蚁、衣蛾或蚊子的时候，也会问自己类似的问题——但没有那么频繁。我的直觉告诉我，在这些情况下，问“谁‘在里面’”这个问题意义并没有那么大。不管怎么说，目睹一只昆虫身体的一部分被碾碎后在地板上痛苦地来回蠕动，总会引发某些灵魂的拷问。

说真的，我提起这些可怖意象的理由，并不是要为一项可能已被大多数读者反复考虑过的事业而摇旗呐喊；相反，我是想提出那个紧要的命题，即一个“灵魂”终究为何物，以及何人或者何物才拥有灵魂。这个命题贯穿在每个人的一生之中——至少也会内隐于心，而对很多人来说

都是十分显在的——也是本书的核心议题。

给我找来一些灵魂更加结实的人

我在前文已经暗示过我对于肖邦音乐深深的挚爱。在少年和青年时期，我在钢琴上弹奏过大量的肖邦，通常都是照着那本由纽约市G.舍尔莫公司（G. Schirmer）出版的亮黄色乐谱版本。打开每一册乐谱，都可以读到一篇由20世纪早期的美国批评家詹姆斯·亨内克（James Huneker）写作的随笔。今天，很多人都觉得亨内克的散文过于恣肆浮夸，但我不这么觉得；他的文字中不加节制的情绪，与我对肖邦音乐的认知发生了共鸣，我也至今尤爱他的写作风格以及他笔下丰富的隐喻。亨内克在他为肖邦练习曲这一册所作的序言中(注)，就《A小调25号第十一练习曲》（这是一次篇幅宏大的爆发，常被人称为“冬风练习曲”，但这自然不是肖邦本人所起的标题，也不是他心目中所设想的意象）做出断言：“灵魂渺小之人，无论他们的手指有多么灵敏，也只能望洋兴叹。”

我本人可以做证，这首汹涌澎湃的乐曲具有令人发指的技术难度。我在16岁左右曾英勇地尝试过学习弹奏这首曲子，可是在中途就悲壮地被迫放弃了，因为仅仅以全速演奏完第一页（在经过几个星期难以想象的艰苦练习之后我才终于设法做到这一点），已经让我的右手一阵一阵地跳痛了。但是，亨内克所指的当然不是这种技术上的难度。他说这首曲子恢宏而高贵，这一点合情合理，可他又更具争议性地在人类灵魂的不同水平或“尺寸”之间划出了一道分隔线，提出有些人就是无法胜任这首曲子的弹奏，这不是因为他们的身体有任何物理上的局限，而是因为他们的灵魂不“够大”。（我不想在这里多费精力去批判亨内克言语中夹带的性别主义；这在那个年代里是司空见惯的。）

这种情感在如今平等主义盛行的美国并不受待见。它在皮奥里亚地

区是行不通的。坦白地说，它在我们这些现代民主社会中听起来太过精英化了，也许甚至到了令人生厌的地步。但我还是不得不承认，我多多少少还是赞同亨内克的，而我也会不由自主地怀疑，即便我们拒不承认，可是在我们每个人深隐的内心里，都依然相信某种隐约类似于人类拥有“小灵魂”与“大灵魂”这种观点。事实上，我忍不住要在此指出，不管我们公开表明自己是多么崇尚平等主义，这种观点确实几乎就是我们所有人共持的一种信念。

小灵魂和大灵魂的人

我们之中有些人相信死刑——刻意地公开扼杀一个人的灵魂，而不论那个灵魂在穿过长长的走廊而被引向处决之地时，多么激烈地乞求宽恕、瑟瑟发抖、颤动、嘶叫或者绝望地挣扎。

我们之中有些人，又或许几乎是所有的人，都相信在战争中杀死敌军的士兵是合情合理的，仿佛战争是一个特殊的情境，缩小了敌人灵魂的尺寸。

在过去，我们之中有些人也许还曾（如乔治·华盛顿、托马斯·杰斐逊和本杰明·富兰克林，至少在某段时期，以他们自己的方式）相信，占有与买卖奴隶并不问情由地拆散他们的家庭，并非不道德之事，就好像我们今天对待猫、狗和马驹一样。

有些信仰宗教的人相信，无神论者、不可知论者和其他信仰的追随者——以及不能更糟糕的是，抛弃了“这种”信仰的叛教者——根本就没有灵魂，因此格外地死不足惜。

有些人（包括某些女人）相信，女人没有灵魂——或者更慷慨一点地说，女人与男人相比，拥有“更小的灵魂”。

我们之中有些人（包括我自己）相信，晚年的里根总统，早在身体停止运转的很多年前，从本质上说已经“消逝”了。由此推而广之，我们也相信阿尔茨海默病末期的病人从本质上说也已经消逝了。令我们深感惊讶的是，虽然这些头颅状的硬壳内仍然盛有一颗人类的头脑，但却有些东西已经从那颗头脑中逝去了——那是某些本质化的东西，其中蕴含着那个人的灵魂的秘密。那个“我”即使并非全部凭空消散，也是部分地消失不见了，逝者如斯，东流而去，再也找不回来了。

我们之中有些人（再算我一个）相信，无论是刚刚受精的卵子还是五个月大的胎儿，都并不具有一个完整的人类灵魂，而从某种意义上来说，妊娠期母亲的生命比她肚子里那个小生灵的生命更为重要，虽然后者毫无疑问也是活着的。

巧克力色拉布拉多猎犬哈蒂

凯莉 吃完早午餐，我们去看林恩的火鸡，我们还从来没看过那东西呢。

道格 那<u>东西</u>，还是<u>那谁</u>？

凯莉 <u>那东西吧</u>，要我说。火鸡不算是一个<u>那谁</u>。

道格 我明白了……那么哈蒂是一个<u>那谁</u>，还是一个<u>那东西</u>呢？

凯莉 噢，她是个<u>那谁</u>，这没问题。

金毛猎犬奥利

道格 对了，今天下午在葛瑞菲湖（Lake Griffy）的郊游，奥利可还

喜欢？

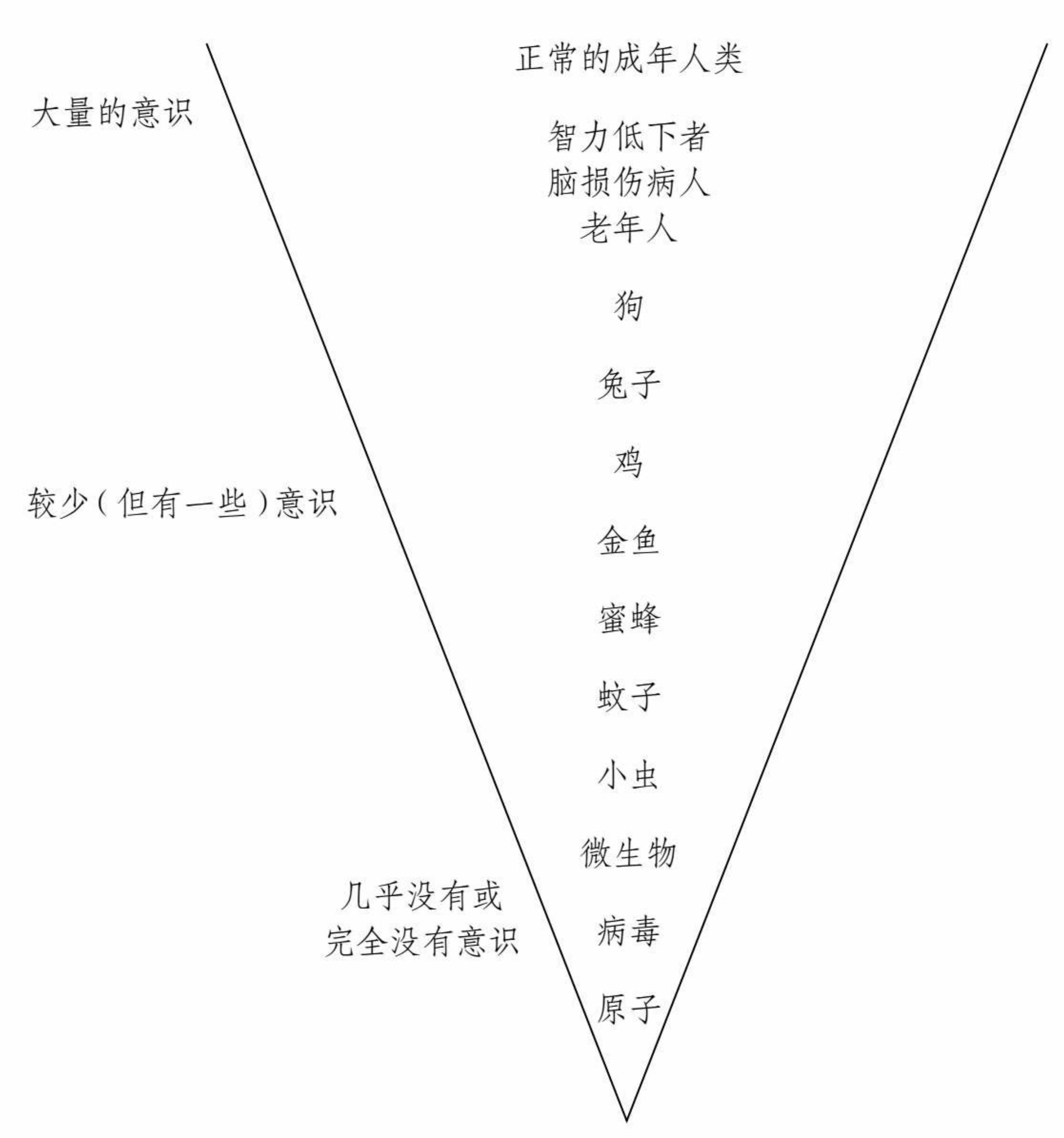
丹妮　噢，他挺开心的，但是他不怎么跟其他的狗狗玩耍。他倒是很喜欢跟人玩耍。

道格 真的吗？怎么会这样？

丹妮 奥利是个有人缘的“人”。

那一条定命而致命的线，该画在何处？

所有的人类，至少所有拥有足够大的灵魂的人类，在拍死蚊子或苍蝇的时候，在放置老鼠夹子的时候，在吃下兔子、龙虾、火鸡、猪乃至狗或马的时候，在购买貂皮围巾或象牙雕像的时候，在使用皮质手提箱或鳄鱼皮腰带的时候，以及甚至于在用青霉素攻击侵入他们身体的细菌群等等时候，都不得不在这件事上做出决定。这个世界时时刻刻都把我们置于大大小小的道德困境之中——最少也包括了每一顿餐食，我们都被迫要选定一个立场。一只羔羊是否也拥有不容轻视的灵魂呢？还是说，羊排就是太过美味，令人无心多此一扰？一条鳟鱼咬上了钩，正在尼龙线的一端无助地挣扎摇摆，它有逃过此劫而幸存下来的理由吗？还是说，我们应该干脆狠狠地给它的头部一记重击，“结束它的痛苦”，这样便可以品味它那种不可名状却偏偏在预料之中的柔软而绵薄的白肌纹质？不管多么黯淡，蟋蟀、蚊子乃及细菌是否也拥有微弱的内在“亮光”？还是说，“在那里面”是完全的黑暗？（在哪个里面？）我为什么不吃狗肉？提供我早餐时享用的那片培根的猪是谁？我在用力咀嚼的是哪一个西红柿？我们应该砍倒前院那株枝繁叶茂的榆树吗？而我在砍树时，还要拔掉那丛野生黑莓灌木吗？贴着它疯长的那些野草呢？



我们这些说话的人凭什么有权㊟决定事关其他那些不会说话的生灵的生死大事？我们（至少对于我们之中的某些人）为什么会觉得自己所处的位置如此令人苦恼？归根结底，这不过是因为强权即公理，我们人类拥有复杂的头脑，并深深根植于丰富的语言和文化之中，得益于由此获得的智力，我们人类相对于“低等”的动物（以及植物）而言，的确更加高级和强大。既然拥有了强力，我们便被迫去建立某种生物的秩序，

而不论我们如此去做是出于长期而审慎的个人思索，还是仅仅在追随不可抗拒的大众潮流。杀死一头牛会和拍死一只蚊子一样心安理得吗？拍死一只在墙壁上招摇过市的苍蝇，会比扭断一只抖成一团的小鸡的头，心里更舒服一点吗？

这类问题显然可以无穷无尽地扩增（proliferate，请注意“扩增”这个动词的反讽意味），但我想就此打住。

我在上面给出了自己的“意识倒金字塔”（consciousness cone）。它并不特别精确，而仅仅是一种示意；但我要提出的是，在你的头脑里，以及在每一个天然具备语言能力的人类的头脑里，都存在某种相似的结构，只不过在大多数的情况下，它很少会受到严格的审查，因为它甚至都不会被明确地阐述出来。

内在性——何者拥有，程度如何？

本书的读者不太可能连一部《星球大战》电影都没看过，也几乎不可能不记得其中颇令人难忘的两个角色——C-3PO和R2-D2。这两个机器人完全是荒诞的空想；像我这样的人，花了几十年的时间在努力通过构建计算模型来理解人类智力最为原初的机制，回头再看这两位，尤其觉得荒唐。

可尽管如此，它们却发挥了一种十分有益的效用——它们开拓了人们的思维。在屏幕上看到了“有血有肉”的C-3PO和R2-D2之后，我们意识到自己并没有一种天生固化的判断程式，让我们每当看到一个由金属或塑料构成的实体时，就会反射性地妄下这种武断的结论：“因为它是由‘错误的材料’构成的[注]，那一定是个无生命的东西。”恰恰相反，我们发现，原来我们可以很轻易地想象出一种由冰冷、僵硬、不似血肉的物质构成的有思考和感觉的实体，这个发现或许会让我们自己大吃一

惊。

我还记得自己在《星球大战》某一部电影中，看到过一支由成百上千整齐划一的行军机器人组成的庞大军队——我所谓的“整齐划一”是指真正的整齐划一，所有的步点都完美地同步一致，而且它们全都挂着一模一样的面部表情，冷漠、茫然、机械。这幅画面明确无误展现出一种绝对的可互换性，我据此猜测，当一颗炸弹降落在这支冲锋之旅的头顶而其全部成员——这些工厂制造的“生物”——在一瞬间被炸成碎片之时，几乎不会有任何观众会为此感觉到一丁点儿悲哀的揪心。毕竟，那些机器人与C-3PO和R2-D2截然相异，根本就不是生物——它们只是大块的金属罢了！这些金属躯壳的内在性，并不比一个开罐器、一辆车或者一艘战舰更多，这个事实由它们完美的一致性暴露了出来。又或者说，即便它们的体内侥幸拥有某些微小程度的内在性，那也跟一只蚂蚁的内在性同属一个量级而已。那些金属的行进军仅仅是机器兵，是属于某个更大的机器人种群中类似雄蜂阶层的成员，而且仅仅是在以它们那种僵尸般的方式，执行植入在它们电路之中的固定的机械化驱动程序。如果真的在某处有些内在性，那也属于足以忽略不计的水平。

那么，到底是什么使我们产生了那种不可抗拒的感觉，相信C-3PO和R2-D2的内在拥有“亮光”，以及在他们无机材料的躯壳里和在他们滑稽转动的“眼”球背后的某处，存在着某些货真价实的内在性呢？我们有关他们的“我”的不可否认的感觉，究竟从何而来？反过来说，前面提到过的晚年里根总统以及被大批炸飞的一模一样的士兵机器人，其内在缺少的是什么？或者说，在巧克力色拉布拉多猎犬哈蒂和机器人R2-D2的内在，到底没有缺少什么，才让我们产生了如此不同的感觉？

一个灵魂的逐渐生长

我在前文中业已陈明，我和很多人一样拒绝承认从人类的精子和卵

子融合成为一个受精卵的那一刻起，一个成熟的人类灵魂便随即诞生了。

恰恰相反，我相信一个人的灵魂是在历经数年的发展过程中慢慢地成长起来的。顺便插说一句："灵魂"一词常常充满着宗教意蕴，在此却丝毫不带任何宗教的色彩；而厘清我赋予这个多变而油滑之词的意义，正是我在本书中的目的所在。我下面的说法听起来可能略显粗糙，但我想至少从隐喻的层面上提出一个"灵魂度"的数值标尺。我们可以首先想象它的范围是从0到100，而这个标尺的单位则可以纯粹为了好玩而被定义为"亨内克"（huneker）。这样一来，我和你，我亲爱的读者，我们两人便都拥有100亨内克的灵魂度，或者差不多这么多。握个手吧！

哎哟！我刚刚意识到自己犯了一个错误，这都怪我常年以来被美国的那种可敬可佩的平等主义传统给洗脑了——也就是说，我在无意识中假设存在一个灵魂度可以"刷爆"的顶值，于是所有正常的成年人在抵达这个天花板之后，就不能向更高发展了。但是，为什么我要做出这样的假设呢？为什么灵魂不能像身高一样呢？成年人群体有一个平均的身高值，但由那个平均值向外延伸的范围是相当广的。成年人的灵魂度为什么也不能同样有一个平均值（比如，100亨内克），外加一个范围宽广的变动区间呢？也许（像智商一样）在极少数的情况下会达到150或200亨内克，而在相反的方向上也会下降到50或更低。

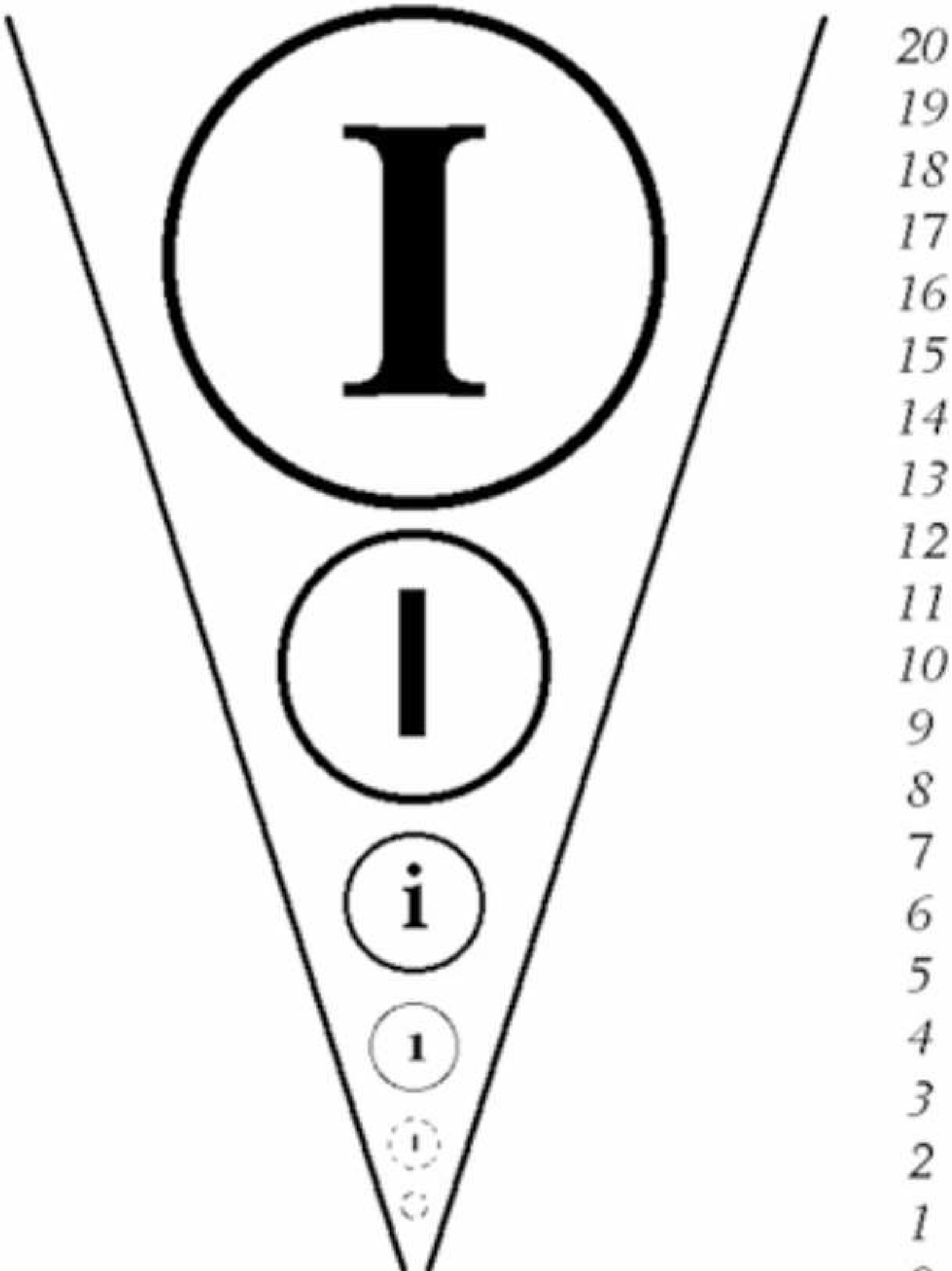
如果事情真是这样，那么我就要收回自己刚才那句不假思索的判断了，即我和你，我亲爱的读者，共同享有100亨内克的灵魂度。我愿意转而提出，咱们两个人在亨内克表上的读数都要比100还要*更高*不少。（我希望你也能赞同。）然而，这就开始感觉有点像是一个危险的道德地界了，近乎是在暗示着一些人比另外一些人*更有价值*——这在我们的社会里是令人深恶痛绝的想法（同时也让我深觉反感），所以我不会在这里浪费很多时间来搞清楚如何计算一个人在亨内克表上的灵魂度数。

让我感到吃惊的是，当精子与卵子结合后，生成的微小生物团所拥

有的灵魂价值实质上是0亨内克。但与此同时发生的是，一个动态的、滚雪球式的实体形成了，它有能力在几年的时间里发展出一套复杂的内部结构或模式——而随着这些错综模式的存在程度越来越高，这个实体（或者不如说是它一步一步慢慢变形成为的那个无比错综复杂的实体群）也在亨内克灵魂量表上获得了越来越高的价值，向100亨内克附近的某一点的价值靠拢。

本页的圆锥图虽然粗略，却生动地勾勒出我可能为0到20岁不同年龄的人们所分配的亨内克值（这也可以代表一个人的不同年龄）。

简言之，我在此是要应和并泛化詹姆斯·亨内克的激进论调，主张“灵魂性”绝不是一个非开即关、非黑即白的离散变量，不像一个比特、一个像素或一个灯泡那样，只有两种可能的状态，而是一个渐变的、边界模糊的数字变量，其值域连续覆盖不同的物种与事物。而我们正在讨论的那种具有特殊类型的微妙模式的实体（阐明其本质占据了本书的大部分工作内容）范畴内，这个变量还可以作为生长与衰亡的结果，随着时间上升和下降。我还要进一步指出，绝大多数的人，不论其承认与否，当涉及是否要吃这种或那种食物、是否要买这件或那件衣物、是否要拍死这只或那只昆虫、是否要支持一部科幻电影中的这号或那号机器人、是否当一部电影或小说中的人物角色死于非命时要为其感到悲伤、是否要宣布某个特定的衰老的人“业已不在”以及种种诸般情景时，他们大体上属于无意识的偏好，正精确地反映了这种存在于思维之中的数字连续体。



20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0

我画的那张圆锥图，在一个给定的人类生命的发展过程中，毫无悔意地标画出了不同的“灵魂度”。读者也许会纳闷，这是不是意味着，当被置于一个巨大的压力情境之下时，比如在电影《苏菲的抉择》（*Sophie’s Choice*）中那样，我会更愿意夺取一个2岁孩子的生命，而放过一名20岁的成人。答案是：“不，并没有这个意思。”即便我真心相信，20岁的成人比2岁的孩子拥有多得多的灵魂（这个观点显然会令很多读者心生不满），那我也会对2岁孩子所具有的*潜力*表示出巨大的敬意，相信他们在随后10多年左右的时间里，可能*发展*出一个更大的灵魂。除此之外，在几十亿年进化机制的打造下，我得以在2岁儿童身上观察到一种特别的东西，因为没有更好的说法，我就将其称作“可爱”好了。这种可被观察到的特质的在场，为2岁孩子提供了一层强大到惊人的坚固保护壳，不仅能抵御来自我的攻击，还能防御不论性别、年龄和信仰如何的全部人类。

点亮了光？

本书的核心目的是尽量精准地敲定那种“别样的微妙模式”的本质。我已经开始相信，正是这种模式奠基或引生了我一直所谓“灵魂”或“我”之物。我完全可以把后者说成是“内在有光”、“拥有内在性”或者作为万年备胎的“有意识”，差别都不大。

心灵哲学家经常使用的术语[注]是“意向性”（intentionality），意思是持有信念、欲望、恐惧等，或“具备语义学”，即具备真正*关于*事物进行思考的能力，而不是玩弄模式复杂的无意义符号的“有限”能力——我在杜撰的苏格拉底与柏拉图的对话中提出过这个区分。

虽然在我们关心的那个令人难以捉摸的抽象概念上，这些术语各自聚焦的方面略有差异，但是从我的视角出发，它们差不多是可以互相替换的。而且我要再次重申，所有这些术语都必须被理解为滑动在一个标

尺上的程度，而不是或开或关、非黑即白、是非分明的开关。

附言

本章的第一稿写成于两年之前，虽然也讨论了肉食主义和素食主义，但是与这个最终的版本相比，则在这个话题上的投入相差甚远。过了几个月之后，我为了“充实文本”，而对《猪》这则短篇小说进行了概述，结果突然发现，我对自己在20多年小心翼翼地划定并自此一直遵循（虽然偶尔会觉得有哪里不对劲）的那条——在哺乳动物与其他动物之间的——分割线产生了怀疑。

突然之间，我对于吃鸡和吃鱼的想法产生了一种格外鲜明的不适感。即便我在过去20多年里都在这么做，还是在震惊之中如“突然戒毒法”一样，戛然而止了。无比巧合的是，我的两个孩子在几乎完全同样的时刻，分别独立地得出了同样的结论，所以在大约几周的时间里，我们家庭的食谱便改换成了彻底的素食。我又返回了自己21岁在撒丁岛时的所在，而这也是我计划坚持永驻之地。

因此，这一章的写作为它的作者带来了一种纯属意料之外的反噬效果——而正如我们将在随后的章节中所见，一个人刚刚做好的选择出乎意料地反弹回来，而这种反转又随之纳入这个人的自我模型，这正是“我是个怪圈”这句格言之含义的最为出色的范例。

---

1. 带着某种复杂的心情，我给母亲读了本书开头的这段话。她虽然已是耋耄之人，但对周围的世界仍然好奇。在一字一句地听我读完后，她感慨道：“如今我发生了很多转变，因为现在对我而言，那些照片就是全部。”无论母亲是不是受了我的影响，我对她能有这样的感慨而感到高兴。
2. On the other hand, [Rucker] proposes that tomatoes, potatoes, cabbages, quarks, and sealing-wax are all conscious.

3. 现在已经不是这样了，请见这一章的附言。

4. Found in [Dahl].

5. 这是伏尔泰《老实人》中潘葛洛斯（Pangloss）的反写，字面意思是“废话”，以示讽刺。

6. All the prefaces that Huneker wrote in the Schirmer editions can be found in [Huneker].

7. See [Singer and Mason].

8. 大脑是由“正确的材料”构成的，而电脑则不是。这是约翰·塞尔的一个口号。

9. See, for example, [Dennett 1987].

第2章
摇曳在恐惧与梦想之间的那只电灯泡

何为“脑结构”？

当人们听说我的研究等同于求索人类思想背后的隐藏机制时，他们总会问我：“噢，这个意思是不是说，你在研究大脑？”

有一部分的我想要回答说：“不，不——我思考的是思考。我思考的是，概念和词汇如何联系在一起，‘用法语思考’是什么意思，是什么造成了口误和其他类型的失误，一起事件如何能不费吹灰之力地引起我们联想到另一起事件，我们如何识别书写的字母和词语，我们如何理解吐字不清、发音含糊和俚语式的言语，我们如何来者不拒地接受那不计其数看似完全平常却史无前例的类比、并偶尔发明出一些灵光闪现的新类比，我们的每一种概念又是如何在一生的时间里于细微处和流动性上不断生长，不一而足。我丝毫不做关于大脑的思考——我把大脑那复杂的、湿乎乎、乱糟糟的网络留给了神经生理学家。”

然而，另一部分的我却想回答说，“我思考的当然是人类的大脑。既然人脑正是执行人类思考的机关，那么从本质上来说，我思考的就是人脑。”

这种耐人寻味的矛盾迫使我拷问自己，“在说起‘大脑研究’时，我指的是什么(注)，而其他人指的是什么？”这又自然引出了下一个问题：“从原则上讲，大脑中的什么结构是一个人可以去研究的？”大多数的神经科学家如果被问起这个问题，都会列出一个包括了（至少一部分）以下

项目的清单（大致按照物理大小排序）：

氨基酸

神经递质

DNA和RNA

突触

树突

神经元

赫布神经集合

视觉皮层功能柱

视觉皮层19区

完整的视觉皮层

左脑

虽然这些全都是神经学合理而重要的对象[注]，但是在我眼中，这个清单却暴露了一种局限化的视角。说研究大脑局限在对如上这些物理实体的研究之内，就好像是在说，文学批评必须聚焦在纸张与装订、墨水及其化学性质、纸页尺寸和页边宽度、字体与段落长度等等之上。可是那些作为文学之心脏的高度的抽象概念又该置之何地呢？——情节与人物、文体与视角、反讽与幽默、典故与隐喻、同情与距离等等。在文学批评家的话题清单上，这些关键的要素消失在何处？

我的观点很简单：不论是在关于文学还是在关于大脑的研究中，抽

象概念才是核心㊟。据此，我附送一份自己提出的“大脑的研究者们”应该同样加以关注的抽象概念的清单：

“狗”的概念

“狗”和“吠”这两个概念之间的联结关系

客体档案（如安妮·特丽斯曼所提出）

框架（如马尔文·明斯基所提出）

记忆组织包（如罗杰·尚克所提出）

长时记忆和短时记忆

情景记忆和旋律记忆

类比桥（如我自己的研究小组所提出）

心理空间（如法康尼尔所提出）

迷因（如理查德·道金斯所提出）

本我、自我和超我（如弗洛伊德所提出）

一个人母语的语法

幽默感

“我”

我可以任意地延长这份清单。它起到的只是一种演示性作用，意在传达出我的论点，即“大脑结构”这个术语应该包括泛属于这一类的项目。不用说，上面这份清单中的一些理论概念不太可能具有持久的有效

性，而另一些则或许会逐渐在不同类型的研究中得到证实。作为一种隐形实体的“基因”[注]使父母的特质得以传递给后代，这一概念的提出与相关科学研究的开展，远在任何物体可以被认定为这种特质的载体之前；作为一切物体之构件的“原子”[注]，其概念的提出与相关科学研究的开展，也远在分离出单个原子并对其进行内部探测之前；同理，上面所列的每一种概念，也许都应被合理地视为一种人脑内部的隐形结构，留待大脑研究者们去进行物理上的厘定。

虽然我也深信，在“人脑”中准确地找出任何一种此类结构（只有一种吗？）的物理载体，都是向前迈步出的令人惊叹的一大步，但我还是不明白，为什么寻找物理图谱就应该构成神经学探究的全部任务。在这种物理鉴别之前（或之后），于上文所列的诸般实体之间构建起各种各样的精确关系，为什么不能也成为大脑研究同样有效的关注对象呢？有关基因和原子的科学研究，在基因和原子被确认为物体及其内在结构得到探测之前，正是如此开展了几十年。

心与脑之间的一个简单类比

我想为关于大脑的研究与关于心脏的研究提供一个简单却至为关键的类比。在我们这个时代，所有人都理所当然地认为，身体及其器官是由细胞所组成的。这么说来，一颗心就是由数十亿个细胞所组成的。专注于这种微观水平上的心脏显然是很重要的，可也有风险因此而错过全景，即心脏是一个泵。类比而言，大脑就是一台思考机器，而如果我们的兴趣在于理解思考为何物，那么我们就不想只见树木、不见森林，甚或一叶障目！只有当我们关注脑的大体建构而不再对其构件展开越来越细微的分析的时候，那幅全景才能清晰地展现。

在10亿年左右之前的某个时间点上，自然选择踩着其惯有的随机脚步，撞见了有节律收缩的细胞。拥有这些细胞的小生命们，自己活得很

不错，因为这种细胞的收缩帮助了有用物质在生命内部的四处传送。如此，偶然之间，泵诞生了；而在所有这种原型泵的抽象设计空间之中，自然选择看好的是更有效的设计方案。

实质上，被发现的是组成这些泵的搏动细胞的内在工作原理，于是，这些细胞的内部构造便不再是获选的关键变量了。这是一场全新的游戏，彼此拮抗的心脏建构成为自然选择的主要竞争者。而在这种新的水平之上，愈加复杂的模式迅速演化开来。

正因如此，心外科医生并不思考心脏细胞的细节，而是把注意力都集中在心脏里大的总体结构之上，正如买车的人不会思考质子与中子的物理性质或合金的化学性质，而是聚焦于高度的抽象概念之上，如舒适感、安全性、热效率、可操控性以及性感指数等等。由是，可为我这一“心-脑类比”画出一条简明的底线以为收束，即如果我们是在寻求解释如概念、观念、原型、陈规、类比、抽象、记忆、遗忘、混淆、比较、创造、意识、同情、共情以及种种极其抽象之现象，那么大脑的微观层面便很可能是——或者不如说几乎肯定是——错误的观察层面。

卫生纸能思考吗？

虽然这个类比相当简单，可令人难过的是，它的底线似乎与很多哲学家、大脑研究者、心理学家和其他对心脑关系感兴趣之人擦肩而过。我们不妨仔细讨论一下哲学家约翰·塞尔（John Searle）的案例。他将自己大部分的职业生涯都花在了大肆挖苦关于人工智能的研究与关于思考的计算模型之上，而且对于嘲笑图灵机特别乐在其中。

一句简短的题外话……图灵机是极端简单的理想化计算机[注]，其存储器由一条无限长（即可以任意延伸）的“纸带”组成，纸带上每一个所谓的“格子”只是一个方块，要么是空白的，要么其中有一个点。一台图

灵机配有一个可动的“读写头”，每次可以照看一个格子之上，可以对其进行“读”（即判断其是否含有点）和“写”（即在其中录入一个点或者擦掉一个点）。同时，一台图灵机还有一份固定的指令清单，存储在它的“读写头”里，告知它应该在何种条件下向左或向右移动一个格子以及录入一个新点或擦掉一个旧点。虽然所有图灵机的基本操作都相当烦琐，但是任何一种计算都能够由一台合适的图灵机来实现（数字是由彼此相邻的可填充点的格子来呈现的，所以两侧为空白格子的“…”就代表整数3）。

言归正传，说回哲学家约翰·塞尔。图灵机是一台抽象的机器，因此原则上可以用任何一种材料制成；塞尔正是对这一事实进行了充分的利用。在我看来，他的花招应该只能糊弄糊弄三年级的小学生，可令人遗憾的是，它竟蒙倒了大批的专业同人：他毫不留情地取笑了这样一种想法，即认为执行思考的系统可能由不着边际的物理材料构成，如卫生纸和鹅卵石（那条纸带可以是一卷无限长的卫生纸，而卫生纸每个方块上的鹅卵石则能扮演格子上的点）、组装玩具或者聚集成一大堆的啤酒罐和乒乓球。

塞尔在其生动的写作中营造出(注)了一种于轻描淡写之间顺其自然地抛出那些幽默意象的表象，可事实上他却是早有预谋，小心翼翼地为他的读者灌输着一种深刻的偏见，又或者，他也许仅仅是在从一种先在的偏见中谋利。毕竟，“思考的卫生纸”这一提法听起来的确荒诞不经（不管那卷卫生纸有多长，也不论是否有鹅卵石这个额外补充），“思考的啤酒罐”“思考的组装玩具”等诸般提法也不例外。这些轻描淡写又似乎顺乎自然的意象，由塞尔举出，以资一哂，实际上塞尔经过了精心的谋算，旨在令其读者对这些想法嗤之以鼻，而不加更多的思考——遗憾的是，它们常常颇为奏效。

口渴难耐的啤酒罐

真的，塞尔对于他以自己那种幽默风格描画出来的系统所引发的嘲弄，实可谓不遗余力。例如，他为了嘲弄由交互联系的啤酒罐组成的庞大系统可能“拥有经验”（这又是一个可以表示意识的词汇）这一想法，举出了口渴这种经验进行讨论，然后似乎好像是漫不经心地征用了一个在每个人看来都十分显见的引喻。他抛出的观念是，在那样一种系统里，必须存在可以“跃然而出”[注]（不用管这到底是什么意思，反正他已经便利地略过了一切有关这些啤酒罐彼此如何交互的描述）的特别的一罐，上面写着“我渴了”几个字。这只单个的啤酒罐（其为一个庞大系统的微观元素，因而可以说相当于大脑中的一个神经元或一个突触）跃然而出，旨在构成系统的口渴经验。事实上，塞尔对这个愚蠢意象的选择是别有用心的，因为他心里清楚，没有人会认为它具有哪怕一点点的可行性。一个金属啤酒罐怎么可能体验口渴呢？而它的“跃然而出”又何以构成口渴？写在啤酒罐上的“我渴了”这三个字，凭什么应该比涂在一辆沾满厚泥的卡车上的“我想被洗一洗”值得更加严肃的对待？

不幸的事实是，对于旨在理解认知与感觉在思维中如何发生的基于计算机的研究而言，这个意象是所有可能中最为荒唐可笑的一种扭曲。批评它的角度俯拾皆是，但我想在此集中分析的一个关键性障眼法是，塞尔如何能漫不经心地宣称，那个啤酒罐大脑模型所具有的经验只局限于一个啤酒罐之上，而他又是如何谨小慎微地避免给出任何暗示，不去提醒人们或许可以在啤酒罐的构造里，于某种更复杂、更整体化和更高层次的属性中，去寻找这个系统的口渴经验。

如果一个人摆正态度，真正严肃地去努力思考一种关于思考或感觉的啤酒罐模型如何能够得以实现，那么就会发现，这种“思考”或“感觉”不论有多么肤浅，都不可能是一种局限在某只单个啤酒罐上的现象。它们须是包含了数百万、数十亿乃至数千亿啤酒罐的庞大过程，而“体验口渴”的状态也不该寄身于事先印在一个可以跃然而出的啤酒罐一侧的三个字之内；恰恰相反，它存在于一种错综复杂的模式之内，涉及数目巨大的啤酒罐。简言之，塞尔讥讽的只不过他自己发明出来的一

个微不足道的标靶。每种感觉或概念仅对应一个孤单的啤酒罐（或神经元）？从来没有任何一位严肃的思维过程建模者会提出这种观念。因此，塞尔的暗箭远远偏离了靶心。

另外，同样值得注意的是，塞尔的“作为口渴体验者的单个啤酒罐”意象其实不过是神经学里一个很早以前便遭遗弃的观念——“祖母细胞”观念——经过改头换面后的变形重现。根据这种观念，你对于自己祖母的视觉再认，当且仅当你大脑中某个特殊细胞被激活时才会发生；而那个细胞便构成了你的大脑对于你祖母的物理表征。一个祖母细胞和一个口渴的罐子之间存在显著的差异吗？完全没有。话虽如此，因为约翰·塞尔拥有一种绘制抓人眼球的意象之天赋，他这种似是而非的论调对很多专业同人、研究生和业余爱好者都造成了经年不散的严重影响。

我在此无意于细节处批驳塞尔（那会耗掉枯燥乏味的整整一章），而只是想指出，如下这种默许假设的影响有多么广泛：一个大脑最原始的物理成分的层级，也必须是这个大脑最复杂、最难理解的思维属性所处的层级。正如一种矿物的诸多方面（它的密度、它的颜色、它的磁性有无、它的光学反射性、它的热传导性与电传导性、它的弹性、它的热容、声音在其中传播的速度以及等等）的属性，取决于其数十亿个原子成分交互作用并形成高级模式的方式，大脑的思维属性也绝非处于单个小成分的层级，而是居于包含了这些成分的大型抽象模式那一阶层。

如果我们打算在对思维现象的分析上取得哪怕只有一丁点儿的进步，那么就绝对有必要把大脑视为多层级的系统[注]。那些难以捉摸的思维现象，是指知觉、概念、思考、意识、“我”、自由意志，诸如此类。

绞尽脑汁地把一个概念、感觉或记忆（等等）锁定在单个神经元上，是完全不可理喻的。而且，当涉及诸如运用类比或者自发地涌起很久以前的旧景等等思考的面向时，甚至把它们局限在一个更高的结构层级之上，比如大脑皮层中的一个功能柱[注]（这些小型结构包含大约40个

神经元，展现着比单个神经元更加复杂的集体反应）上，也是毫无意义的。

大脑中的层级与作用力

我曾经看到过一本题为《分子上帝：分子如何决定我们的行为》（*Molecular Gods: How Molecules Determine Our Behavior*）的书[注]。虽然我没有把它买下来，但是它的标题在我的大脑里激发了很多想法。（什么是大脑里的一个想法？一个想法真的是在一个大脑的内部吗？一个想法是由分子构成的吗？）确实，我很快就把这本书放回了书架上，这个事实本身就是一个很好的例子，反映着它的标题在我的大脑中所引发的种种想法。准确地说，究竟是什么决定了我在那天的行为（例如，我对那本书的兴趣、我关于它的标题的深思、我不买这本书的决策）？让我把它放回架上的，是在我大脑内部的某些分子吗？或者，那是我大脑中的某些观念吗？当我先翻了翻那本书，然后把它放回去的时候，于我脑海中发生之事，有没有一种合适的谈论方式？

那时候，我正在阅读很多不同的作者讨论大脑的著作，在其中的一本书里，我偶然读到了由神经学家罗杰·斯佩里（Roger Sperry）执笔的一章。那一章的写作不仅饱含一种特别的热情，还表达了一种与我自己的直觉发生了强烈共鸣的观点。我谨在此引用斯佩里的论文《思维、大脑和人道主义价值》中的一小段[注]。我认为，这段文字尤其发人深省。

> 在我自己假设的大脑模型中，意识自觉确实呈现为一个十分真实的因果主体，并在大脑事件的因果关系和控制链条中占据了一个重要的地位，显示为一股积极的、操作性的力量……
>
> 非常简单地说，最终归结到一个问题就是，占据在头盖骨中的诸般因果力量，到底是谁在使唤谁。换句话说，这关乎大脑中控制

成分的等级关系。在大脑内部存在一个由不同因果力量构成的完整世界；除此之外，还存在力量内部的力量之内部的力量，这不同于我们所知的宇宙中任何其他某种半立方英尺大的事物……

长话短说，如果一个人在大脑内部的指令链条上不断向上攀爬，就会在最高处发现那些全局性的组织力量，以及那些与思维状态或心理活动相关的大脑应激反应全局模式的动态属性……

而在大脑中这一指令系统的顶点附近……我们所遇的即是观念。

人类比大猩猩强在拥有观念与理想。在这里提出的大脑模型中，一个观念或一种理想的因果潜力变得同一个分子、一个细胞或一次神经冲动一样真实。观念引发观念，并协助新的观念发生演变。它们彼此之间交互作用，并与其他思维力量交互作用。这些思维力量可能来自同一个大脑内部，也可能来自邻人的大脑，而得益于全球化的交流，还可能来自距离遥远的外国人的大脑。此外，它们还与外部的环境发生交互作用，在进化的过程中形成突破性进展，远远超过任何先在的进化景观，甚至超过了有生细胞的出现。

在头盖骨内部是谁在使唤谁？

没错，我亲爱的读者，我是在问你：在那个纠缠交错的庞大神经系统，亦即你的大脑里，是谁在使唤谁？而在“摇曳在恐惧与梦想之间的那只电灯泡”[注]，亦即我的大脑里，又是谁在使唤谁？

斯佩里对大脑内部等级关系的质询，确切地指出了我们关于自己需要了解之义——或者，更较真儿地说，是关于我们的自我。当我们说起某个自称为“我”之物做了某件所谓“决策”之事，随后与其相接的附属肢体流畅地运动起来，让一本书又回到了几秒钟之前的位置，在那样美好的一天，在那个运转良好的大脑里，真正发生的究竟什么？真的有某种可被称为“我”的东西在“颐指气使”地支配各种不同的大脑物理结构吗？

其结果是通过神经纤维传达了经过周密调和的特定信息，并引发了肩膀、手肘、手腕和手指的相继运动，而这种以某种特定的复杂模式展开的运动，把那本书原封不动地遗留在原处。还是恰恰相反，那仅仅是发生在时空连续体局部的千千万万个微观物理过程（涉及电子、质子、胶子、夸克等等粒子之间的量子机械碰撞），而诗人埃德森正是把这种时空连续体昵称为“明暗不定的灯泡”。

恐惧与梦想、希望与哀伤、观念与信仰、兴趣与疑虑、满足与嫉妒、记忆与野心、阵阵乡愁与恻隐之心、负疚的闪现与天才的火花：它们在物理世界中是否有一席之地？如此纯粹的抽象概念是否具有因果力？它们可以支配实在之物吗？或者仅仅是无效的虚构？一个模糊不清、无影无形的“我”能够给如电子或肌肉（或者就事论事，也可以说书本）等具体的事物下达行动指令吗？

宗教信仰曾经引发过战争吗？还是说，所有的战争都是由数万亿（这已经近乎荒诞地低估了真实的数字）极其微小的粒子遵循物理定律进行交互作用而引发的？是火引发了烟吗？是汽车引发了雾霾吗？是嗡嗡声引发了无聊吗？是玩笑引发了大笑吗？是微笑引发了昏厥吗？是爱情引发了婚姻了吗？或者归根结底，是不是只存在海量的粒子在遵循物理定律相互支配——最终没有给自我或灵魂、恐惧或梦想、爱情或婚姻、微笑或昏厥、玩笑或大笑、嗡嗡声或无聊、汽车或雾霾乃至于烟与火留下任何空间？

热动力学与统计力学

伴我从小长大的父亲是一名物理学家，所以对我而言，将物理学视为潜伏于发生在宇宙中的每一件事的表面之下的基础，是再自然不过的了。即使当我还是个年纪很轻的男孩子时，我就从通俗科普读物中得知，化学反应是交互作用的原子的物理学造成的结果，而当我阅历见长

后，我把分子生物学也看作是物理学定律作用于复杂分子的成效。简言之，在我的成长过程中，我眼中的世界装不下任何“额外的”力量，任其可以超越或凌驾于物理学家们已经认定的四种基本力（重力、电磁力和两种核力——强核力与弱核力）。

可我又是如何做到，随着年纪的增长，于自己坚如磐石的信仰之外接纳了外加的信条？我相信是进化引发了心灵的演变、是宗教教义招致了战争、是怀乡之情启发肖邦创作了一首特定的练习曲、是强烈的专业嫉妒心催生了大量恶意书评的写作，不一而足。这些很容易把握的宏观因果力，似乎从根本上便与四种妙旨难言的物理力迥然而异；而我曾经无比确信，是后者引发了宇宙里的每一起事件。

答案十分简单：我仅仅是把那些“宏观的力量”想象成了由基本物理力所引发的复杂模式的描述方式而已。这很像是物理学家们逐渐发现，如摩擦力、黏性、透明度、压力和温度这类宏观的现象[注]，都可以被理解为由微观成分的统计学所决定的具有高度预测性的规律现象，而这些肉眼难见的微观成分，多于恒河沙数，在时空中横冲直撞、相互碰撞，一切都遵循着四种基本物理力的指令。

我还意识到，这种在描述层级上的转换，为生物提供了某种十分宝贵的东西：可理解性。写一篇包含了阿伏伽德罗常量[注]那么多个方程的长文来描述一种气体的行为（假设这项艰巨无比的壮举是可行的），不会引导任何人对任何事达成理解。

可是，放弃细节信息而制作一份统计概要对于可理解性的增加却是大有裨益的。当我形容“满地秋叶”而不必详细说明每一片的形状、方向与颜色时，我会感到轻松自在；同理，当我提到一种气体时，如果只需描述它的温度、压力和体积而不必详考其他细节，我也同样会感到自在舒心。

诚然，所有这些对于所有的物理学家和绝大多数哲学家而言都是陈

词滥调，而且可以用一句没什么新意的格言概括出来：热力学可由统计力学来解释。但是也许把这个观念倒转过来，它反而会多多少少变得更加清晰一些，这样一来：在热力学层面进行讨论时，统计力学可以被绕过不提。

我们作为动物而存世，知觉局限在宏观的日常事物的世界之内，显而易见，这促使我们在正常的生命运转中，不去参照任何微观层面上的实体与过程。大约100年以前，还没有人对于原子有一丁点儿真正的了解，可人们依然过得很好。麦哲伦完成了环绕地球的航行，莎士比亚写出了数篇戏剧，巴赫创作了若干首套曲，而圣女贞德也焚身于火刑柱之上，他们每个人都有自己的恰当（或不恰当的）理由，但从他们的视角出发，无论是谁，都与DNA、RNA和蛋白质毫无关系，也不涉及碳、氧、氢、氮，或者光子、电子、质子、中子，更别提夸克、胶子、W和Z玻色子(注)、引力子和希格斯粒子了。

思考力学与统计思维学

如此说来，不同的描述层级，依据不同的目的和语境，具有不同种类的功用性，这对任何人来说都已不再是新闻了；而我将这一简单的真理应用于思考与大脑的世界之中，据此总结出了一个自家的观点：思考力学可由统计思维学来解释。同样还有它的倒转版本：在思考力学层面进行讨论时，统计思维学可以被绕过不提。

我所说的这两个术语，“思考力学”和“统计思维学”，是什么意思呢？其实相当直白。思考力学类比于热力学，专注于脑中大规模的结构与模式，完全不涉及如神经放电一类的微观事件。思考动力学是心理学家研究的对象：人们如何做出决策、如何犯下错误、如何感知到模式、如何体验新奇的提醒，诸如此类。

相比之下，我所说的“思维学”指的是小规模的现象，通常是神经学家研究的对象：神经递质如何在突触之间传播、细胞如何连接在一起、细胞集群如何共振，诸如此类。而我使用“统计思维学”指那些非常小的实体经过平均化以后的集体行为——换句话说，是一个巨大蜂群的整体行为，而不是其中的一只小蜜蜂。

不管怎么说，就像神经学家斯佩里在前文所引段落中已经清楚地表明的那样，大脑与气体不同，其中并不存在一条单一的由基本成分直达整体的自然上跃；相反，在从思维学上达思考热力学之间的通路上，经过了很多中间站，而这就意味着我们尤为难以找到，甚或难以想象出，如何在基础层面和神经水平上去解释，为什么某位特定的认知科学专业的教授曾经选择把一本专论大脑的书放回了书架，为什么他有一次忍住没有拍死某一只苍蝇，为什么他曾在一场肃穆的葬礼上忽然爆发了一阵咯咯大笑，又为什么在哀叹一位受人敬爱的同事的离世时高声大呼：“她将永垂不朽！”

每日生活的压力要求并强迫我们在讨论事物时，停留在我们直接感知到它们的层面。我们的感官、语言和文化为我们提供了连接那个层级的通路。从最早的童年时代开始，像“牛奶”“手指”“墙壁”“蚊子”“叮咬”“瘙痒”“拍打”等等概念就被放在银色托盘上呈递给我们。我们经由这些概念来感知世界，而不是通过微观的概念，如“细胞质”、“核糖体”、“肽键”或“碳原子”。我们当然可以在后天习得这些概念，我们之中有些人还会对它们形成深刻的把握，但它们永远也不可能替代那些在银色托盘上伴随我们成长起来的概念。于是，总而言之，我们都是我们的宏观视野的囚徒，逃不出使用日常言语来描述我们目睹并感知为真实的事件的陷阱。

为什么相比于把一场战争努力想象为由交互作用的基本粒子构成的大型模式并以类似方式思考引发战争的动因，我们说一场战争是由宗教或经济的原因所引发，就要自然许多？其原因正在于此。物理学家们或

许会坚持认为，前者才是唯一“真正的”解释层面；其道理在于，如果我们在这个层面上展开讨论，那么任何信息都不会被漏过。但是，哎（或者不如说“谢天谢地！”），持有这种超凡的精准性并非我们的宿命。

我们这些凡人已经被宣判为不可在没有任何信息损失的条件下谈话。我们必须做出简化，而且实际上也做出了大量的简化。但是这种牺牲同时也是我们的荣耀。正是这种极端的简化，令我们得以拨开烦冗，看清状况的本来面目、发现抽象的本质[注]、明确关键的要点、站在惊人的高度上理解现象、可靠地在这个世界上并构想出文学、艺术、音乐与科学。



1. See [Churchland], [Dennett 1978], [Damasio], [Flanagan], [Hart], [Harth], [Penfield], [Pfeiffer], and [Sperry].

2. See [Damasio], [Kufiler and Nicholls], [Wooldridge], and [Penfield and Roberts].

3. See [Treisman], [Minsky 1986], [Schank], [Hofstadter and FARG], [Kanerva], [Fauconnier], [Dawkins], [Blackmore], and [Wheelis] for spellings-out of these abstract ideas.

4. See [Judson].

5. See [Pais 1986], [Pais 1991]), [Hoffmann], and [Pullman].

6. See [Hennie] and [Boolos and Jeffrey].

7. See Chapter 22 of [Hofstadter and Dennett].

8. 塞尔在那篇狂妄自大地诋毁侯世达与丹内特的文章中称："让我们来想象一下，我们那个用来激发口渴感觉的程序，运行它的电脑完全是由旧啤酒罐组成的，数百万乃至亿万个旧啤酒罐被连接在杠杆上，并由风车驱动。我们可以想象，这个程序是通过让啤酒罐互相撞击来刺激突触处的神经激活的，因此在神经激活与啤酒罐之间达到了一种严格的一一对应关系。而在序列的末端，一个啤酒罐会跃然而出，罐子上面印着'我渴了'。下面再回到原来的那个问题，怎么会有人愿意承认，这套鲁布·戈德堡式的装置，真的像你我一样感受到了渴意呢？"

9. See [Simon], [Pattee], [Atlan], [Dennett 1987], [Sperry], [Andersen], [Harth], [Holland 1995], [Holland 1997], and the dialogue "Prelude... Ant Fugue" in [Hofstadter 1979] or in [Hofstadter and Dennett].

10. See [Kuffler and Nicholls].

11. This was [Applewhite].

12. Taken from [Sperry].

13. 引号中那个具有奇妙的唤起性的短语，取自美国诗人拉塞尔·埃德森（Russell Edson）的《地板》（*The Floor*）（See [Edson], which is a thin, remarkably vivid, highly surrealistic, often hilarious, and yet profoundly depressing collection of prose poems.），同时也用作了本章的标题。

14. A beautiful and accessible account of the emergence of everyday phenomena (such as how paper tears) out of the surrealistically weird quantum-mechanical substrate of our world is given in [Chandrasekhar].

15. 热学常量，也与化学中物质的量（摩尔）紧密相关，指0.012kg碳12中包含的碳12的原子的数量，一般取6.02×10^{23}。——编者注

16. See [Pais 1986] and [Weinberg 1992].

17. See [Kanerva], [Kahneman and Miller], [Margolis], [Sander], [Schank], [Hofstadter and FARG], [Minsky 1986], and [Gentner et al.].

第3章 模式的因果潜力

原动力

鉴于本书余下的所有部分都离不开对于思考实体不同描述层级之间相互关系的清晰认识，我在此愿意分享几个具体的隐喻，它们对于我发展出有关这个捉摸不透的主题的直觉，产生过很大的帮助。

我的第一个例子涉及的概念为人熟知，是一连串翻倒的多米诺骨牌。只不过，我要把这个标准的意象变得灵活有趣一点，规定每张多米诺牌中都以一种巧妙的方式装载着弹簧（我们不关心具体的细节），经过一段“顽固”期后，它会弹回原来的竖直状态，而整套牌会被再推倒一次。我们可以用这样一种系统实现一种机械计算机，把信号沿着多米诺骨牌传递下去，那些牌可以分叉，也可以交汇；如此一来，信号既可以在环路中扩增，也可以联合引发其他的信号，执行诸如此类的行为。当然，相对时间的调准将会是成败的关键，但我要再重申一遍，我们所关心的并非细节。这里基本的观念仅仅在于，我们能够想象出一个经过了精确时间调配的多米诺骨牌链阵网络，而其相当于执行某种特定运算的计算机程序，比如判断一个输入的数字是否为质数。（对不同寻常的计算机材料格外钟爱的约翰·塞尔，理应爱上这个“多米诺链阵”思考实验！）

接下来，让我们想象自己为这个链阵提供了某个特定的数字“输入”，选取的是我们所感兴趣的任意一个正整数——比方说，641[注]，然后把正好那么多张多米诺骨牌首尾相连地摆放在那片网络中的一个“保

留”分支之中。现在，当我们推翻链阵上的第一张多米诺骨牌时，就会发生鲁布·戈德堡式（Rube Goldberg-type）的一系列事件：多米诺骨牌一张接着一张地翻倒。启动不久之后，所有构成我们输入分支的641张多米诺骨牌都会倒下，其结果是触发了各种各样不同的环路，某个环路检验输入的数字能不能被2整除，另外一个环路则检验其能否被3整除，依次类推。如果能够找到一个因数，就会有一个信号沿着一条特别的分支传送下去，我们可以为这条分支命名为“因数分支”，而每当看到这条分支倒下，我们就会知道，输入的数字拥有因数，因而并非质数。相反，如果输入值没有因数，那么因数分支就永远不会被触发，而我们就会知道，输入值是一个质数。

设想在多米诺链阵中输入641这个数字之时，一位观察者正站在旁边。这位观察者事先并不知道排出这个链阵的意图何在，于是在用心观察一段时间之后，他指着因数分支中的一张多米诺骨牌充满好奇地发问：“为什么那里的那张多米诺骨牌从来都不倒呢？”

可能会有人为他提供两种完全不同类型的答复，下面让我来对比一下。第一种类型的答复采取了一种微观到愚蠢的视角，是这样的：“因为它前面那张牌从来不倒呀，你这个白痴！”当然，就其涉及的范围而言，这个答案是正确的，可是其有效的范围十分有限。它只是把问题转移到了另一张多米诺骨牌之上，从而回避了问题的实质。

第二种答复是：“因为641是一个质数。”而这个回答虽然也同样是正确无误的（实际上，从某种意义上说，它更加切题），却神奇般地没有提及任何具有物理性质的东西。不仅聚焦点向上移动到了链阵的集体属性之上，而且那些属性以某种方式超越了物理层面，而与纯粹的抽象概念相关，比如说：质数性。

第二种答复绕过了所有关于重力和多米诺骨牌的物理学，只涉及隶属于一个完全不同的论域的概念。质数的论域与翻倒多米诺骨牌的物理学相去甚远，就如同夸克和胶子的物理学与冷战时期认为激进主义将不

可避免地逐个颠覆东南亚国家的“多米诺骨牌理论”一样遥远。在上述两种情况中，两个论域之间相隔了许多层级，其中一个具有纯粹的局部性和物理性，而另一个则是全局化和组织化的。

在转入下一个隐喻之前，我还想最后指出一点：641的质数性，虽然在此被用来解释为什么某张特定的多米诺骨牌没有倒下，但是它也完全同样可以用来解释为什么另一张不同的多米诺骨牌就翻倒了。特别是，在那个多米诺链阵中，还可以有一个命名为“质数分支”的分支，当整套可能的除数全部被穷尽时，这条分支上所有的多米诺骨牌都会倒下，这也就意味着输入的数字被确认为是一个质数了。

这个例子的要点在于，641的质数性是关于为什么某张特定的多米诺骨牌会倒下以及另外一些多米诺骨牌不会倒下的最佳解释，也许甚至是唯一的解释。一言以蔽之，641是主使者。所以我要问的是：在多米诺骨牌链阵内部，是谁在使唤谁？

集体现象的因果效力

我的下一个隐喻是自己在不久前的一天下午凭空捏造的。当时我正开车在乡下的某条高速公路上，被一场可怕的交通堵塞困住了。不同车道上的车与车之间几乎是摩肩接踵，都像石头似的一动不动。不知出于什么原因，我想起了大城市里的交通堵塞。在那里，你常常能听到人们愤怒地向彼此狂按喇叭。于是，我想象自己突然之间开始朝着在我前面的人狂按喇叭，好像在说：“把路给我让开，你这个蠢货！”

想到自己（或者任何人）做出这样一种幼稚到离谱的举动，让我不禁一笑，但是当我稍加考虑，我发现那种鸣笛方式可能真的具有一点微小的合理性。毕竟，如果下一辆车能够“嗖”的一下如魔法般地消失，那么我就能填补那个空位，并前进一辆车的长度。虽然一辆车“嗖”的一下

凭空消失的可能性没那么大，而且前进一辆车的长度也不算多大的进步，但是凭借这个意象，鸣笛的观念以某种方式变得近乎可以让我理解了。然后，我又记起了我的多米诺骨牌链阵和那个愚蠢的超级局部化的回答：“那张多米诺骨牌没有倒下，是因为它相邻的牌没有倒，你这个白痴！”这一短视的答案以及我关于朝前面的车按喇叭的一闪念，看起来如出一辙。

我继续停坐在交通拥堵之中，心不在焉地摆弄手指，而没有按响喇叭，同时任由这些想法颐指气使地摆布着我无助的神经。我想象了一个与事实相反的场景，为高速公路裹上了一层可以想象到的最浓的豌豆汤一样的浓雾，而我几乎看不清在我前面行驶的车辆的尾部。在这种情境下，鸣响喇叭就不会显得那么蠢了。就我所知，也许引发我寸步难行的全部原因就是那辆车，但凡它能让开路线，我就可以在高速公路上畅行无阻了！

如果你是像这样完全地被浓雾笼罩，或者你是个高度的近视眼，那么你就可能对自己想，“这都是我邻车的错！”而且你至少有可能是正确的。但如果你享有一片更为宽阔的视野，并且能够看到四周成群拥挤的一动不动的车辆，那么朝你的前车按喇叭就很荒唐了，因为很明显的是，问题并不在于局部。根本的问题在于汽车以外的某种话语层面上。虽然你也许不了解问题的本质，但是在这场交通堵塞背后，一定潜伏着某种更高层级、更加抽象的原因。

也许一场至关重要的棒球比赛在道路前方3英里的地方刚刚结束。也许这是一个工作日的早晨7点30分，而你前往的目的地是硅谷。也许10英里之外发生了一场巨大的暴风雪。又或者也许是其他一些事情，但其必然是某种诸如此类的社会或自然事件，足以引发很多人做出同样的事情。不管你在汽车机械方面的造诣有多高，都不会有助于你把握这样一种情境的实质；你需要的是有关可以作用在高速公路与交通之上的那些抽象力量的知识。汽车在这盘更大的棋局中，不过是小兵小卒而已，

更不用提它们无法穿越彼此并在穿越后保持完整（就像水波和其他波所做的那样）的事实了。它们的物理本性在交通拥堵中不发挥任何显著的作用。我们所处的这个情境可以与多米诺骨牌链阵的提问场景相类比；在后者当中，“641是个质数”这个全局的、抽象的、数学层面的答案，比一个局部的、物理的、多米诺骨牌层面的答案何止优越百倍。

神经元与多米诺骨牌

前文提到的那些生活中的场景，为我们讨论人脑中因果性的诸多层级提供了有用的隐喻。假设监视我大脑中任意选定的神经元是可能做到之事，在这种情况下，也许会有人提问，在我听到某段音乐时，“为什么45826493842号神经元总是不被激活呢？”一个局部而短视的答案可能是：“因为接入它的那些神经元总是不会联合激活。”这个答案同其他情境中的短视答案一样，正确无误，却既无用，也无效。相反，全局的、组织化的答案则更切中题旨：“因为侯世达不中意法兹·多米诺的曲风。”

当然，我们不应落入思考的陷阱，认为45826493842号神经元是当我与某首正在听的音乐形成共鸣时被激活的独一的神经元。它只是很多参与到这一高级过程中的神经元之一，就像国民大选中的一名投票者一样。正如没有任何一名特别的投票者可以做出决定一样，也没有任何一个神经元是享有特权的。只要我们远离诸如享有特权的“宏观”这种简化概念，我们就可以使用多米诺骨牌链阵这一隐喻来思考大脑，尤其是以此提醒自己，对于大脑中一个给定的现象，何以存在着大量的不同解释，而它们分属各种不同的抽象层级之上的诸多不同论域。

作为原因的模式

我希望在这些意象的协助下，罗杰·斯佩里有关在一个如大脑或链阵这样复杂的系统中“因果力的分布”与“组织的整体效力与动态属性”的论述，已经具有了更加清晰的轮廓。例如，让我们试着回答这个问题：“641的质数性真的可以在一个物理系统中起到一种因果作用吗？”尽管641的质数性显然并非一种物理力量，可答案仍然必得是：“没错，它确实发挥了一种因果作用，因为对于链阵行为最为有效和最具洞见的解释，其关键便系于这个概念之上。”对于因果性的深度理解[注]，有时候要求理解非常大的模式及其抽象的关系与交互作用，而不仅仅是理解微观事物在微小时距中的交互作用。

我不得不强调说，在此并没有“额外的”物理（或物理之外的）力量；局部的、短视的物理学法则自行运转良好，但是多米诺骨牌的全局安排决定了何事将发生，而且如果你注意到（并理解了）这种安排，那么关于在除数分支上没有倒下的多米诺骨牌（以及在质数分支上倒下的多米诺骨牌）的答案，便会有一条载满洞见的捷径，被置于银色托盘之上，呈现在你的面前。反之，如果你没有注意到这种安排，那么你将注定绕上一条远路，只能理解局部之事，而得不到任何洞见。简言之，把641的质数性视为我们多米诺骨牌链阵的一个物理动因，类似于把一种气体的温度作为一种物理动因（例如，造成了其向承载它的容器壁上施加的压力）。

确实，让我们暂停片刻，想一想这样一种气体——它被装在一个配有可移动活塞的气缸内。如果该气体突然之间升温了（当你的汽车引擎中任意一个气缸在塞口迸出火星的时候，就是发生了这种现象），然后它的压力突然增加，因此（注意这个指示因果的单词）活塞也被突然间向外推出。这正是内燃机制造的原理。

我刚刚讲的这个故事处于一个总体（热动力学）的层面之上。没有哪个设计内燃机的人会在意微观的层面，亦即分子的层级。机师们才不会费心去理清10^{23}个分子互相碰撞的精确轨道呢！个体分子的坐标与速

率纯粹是与此无关。重点在于，我们可以指望它们集体向外推动活塞。事实的确如此，至于它们究竟是X类型、Y类型还是Z类型的分子，统统无关紧要——压力就是压力，而这才是全部的重点所在。作为一种高层级的事件，爆炸将会执行加热气体的功能，而气体也将推动活塞。这种对于所发生之事的高层级的描述，是唯一恰当的描述层级，因为所有的微观细节都可以发生改变，但（至少从一位人类机械师的视角出发）完全相同的事情依旧会发生。

更低层级的奇怪的无关性

位于底部的层级，虽然百分之百是造成所发生之事的根由，可是却与所发生之事无关，这个观念听起来几乎是自相矛盾的，但它却是一个日常的自明之理。为了让这一点透彻明晰，我还要再举一个例子来进行阐明。

回想在我八岁的时候，当我初次在父母的留声机上听到《肖邦练习曲第25号》第四首时，我立即爱上了它。现在假设我的母亲把唱针置于槽纹上的时间拖后了0.001秒。可以肯定的是，那个房间里所有的分子都会处在完全不同的状态之中。如果你是那些分子中的一员，那么你的故事将会发生难以想象的改变。由于那0.001秒的延迟，你将在完全不同的地方猛冲并撞击到完全不同的分子群，来自完全不同的方向，如此继续，永无休止。不管你是哪个房间中的哪个分子，你的生命故事都会发展到一个难以想象的不同状态。但是有任何这种不同对那个听音乐的小男孩产生了哪怕一丁点儿的改变吗？没有，连微乎其微的一丁点儿改变也没有。唯一有影响的是，《肖邦练习曲第25号》的第四首要穿过空气忠实地传播出来，而这一点无比肯定必将发生。如果我的母亲把唱针插入槽纹中的时间提前或延后0.001秒，我的生命故事不会有任何形式的改变。或者，早一秒或者晚一秒也无所谓。

虽然空气分子对于一系列涉及某个特定的孩子和一首特定的音乐的高级事件而言，是至关重要的中介因子，但它们具体的精确行为却并非关键所在。确实，说其“并非关键”是一种保守得近乎荒谬的陈述。这些空气分子能够以无数种人类难以辨别的方式，来完成一模一样的关于孩子-音乐的任务。它们相互碰撞的低层级法则所起到的作用，仅仅是导致可以预测的高层级事件的发生（把肖邦练习曲的音符传送到小道吉的耳朵里）。但是这些分子的定位、速度、方向乃至化学特性，所有这些都是可以变换的，而高层级的事件仍可以保持不变。在我的耳朵里听起来，还是一样的音乐。甚至可以想象，连微观的物理法则也可以变动——重要的并不是细节的法则，而仅仅是它们能可靠地导出稳定的统计结果这一事实。

投掷一枚25美分的硬币100万次，你拿到50万次正面的概率将十分可靠地小于1%。投掷一枚10美分的硬币同样的次数，同样的结论依然成立。每次投掷都使用一种不同的硬币—10美分、25美分、新币、旧币、水牛镍币、银币，凡是你能想到的都可以——而你还是会得到同样的结论。把你的10美分硬币从圆形削成六边形的轮廓——结论无差。把六边形的轮廓换成大象的轮廓。在每次投掷之前把硬币在苹果酱蘸一蘸。把投掷的方式改成用一根棒球棒把硬币高高地打向上空。在氦气而不是空气中投掷。在火星而不是地球上做这个实验。以上种种以及其他无数种具体的变量都不会对事实产生任何影响，即在100万次投掷中，50万次正面朝上的概率不足1%。这种高层级的统计结果是稳健的，不会随着基底的细节或掌管投掷与反弹的微观法则而变化；这种高层级的结果是绝缘的，封装在微观层面之外。它本身就是其自己所处层级之上的事实。

我说过，虽然在较低层级上发生之事是造成在较高层级上发生之事的根由，但是却仍然与较高层级不相关，此话的意义正在于此。更高的层级可以无牵无挂地忽视更低层级上的过程。正如我在第2章中所言：“我们作为动物而存世，知觉局限在宏观的日常事物的世界之内，

显而易见，这促使我们在正常的生命运转中，不去参照任何微观层面上的实体与过程。直到大约100年以前，还没有人对于原子有一丁点儿真正的了解，可人们依然过得很好。”

向不可预见性的波谱脱帽致意

我并未暗示那种隐形的、密集的、混乱而微观的世界层级可以被彻底地抹去与遗忘。虽然在很多情况下，我们都指望自己熟悉的那个宏观世界完全可以被我们预见，但是也存在很多其他状况，当我们身处其中之时，十分清楚自己没有能力预见接下来将要发生之事。不管怎么说，首先让我列举一张简短的可以预见的样本清单；一直以来，我们都在不假思索地依赖着它们。

当我们转动汽车的方向盘时，我们确信地知道车将驶向何方；我们不必担心一队拒不服从的小分子可能会发动叛乱，并破坏我们的转向。当我们把装满水的炖锅底下的一个电火炉调到“高”挡时，我们知道锅中的水会在几分钟之内沸腾。我们无法预测沸水中的气泡翻腾的模式，但是我们对此真的毫不在乎。当我们从杂货店架子上取下一个汤罐头并把它放在我们的购物车中时，我们确定地知道它不会变成一袋薯片，不会烧伤我们的手，不会沉重到我们无法举起它来，不会从购物车的条栏之间脱漏，而且只要把它摆正，它就会一动不动地安坐在车中，如此种种。可以肯定的是，如果我们把汤罐头横过来放置，然后开始推着购物车在店里走来走去，这个罐头就会在车里以我们不可预测的路径滚来滚去，但是这些路径也完全不出我们可以想见的范围，而且对于我们来说，这除了有点轻微的恼人之外，既引不起我们的兴趣，也没有什么重要性。

当我们开口说话的时候，我们知道这些言语会抵达我们听众的耳朵，不会由媒介的压力改变成另外的话，而且甚至还会精确地贯彻我们

加诸它们之上的语调。当我们把牛奶倒进一个玻璃杯里时，我们知道该把牛奶盒倾斜多少，才能获得我们想要的流量，从而避免溅出一滴牛奶。我们控制牛奶，并达到正好为我们所希望的结果。

上述任何一种情况，都毫无惊奇可言！我可以无限地延长这份清单，而它很快就会变得异常乏味，因为你本能地知道所有这些事，而且完全将其视为理所当然。在生活的每一天里，我们都在以100万种心照不宣的方式，在那个看得见也摸得着的世界之中，依傍着关于事物如何发生的可预见性；而这些可预见性不计其数，如磐石一般确凿（磐石的坚硬又是这些如磐石一般确凿的可预见性之一例）。

另一方面，在“上面这里”的宏观世界中，也存在着大量的不可预见性。再来一份列举典型的不可预见性的清单，如何？

当我们把篮球投向篮筐的时候，我们不知道会不会投中。这一球也许砸到了篮板，然后弹到篮筐边沿，又转悠了几秒钟，让我们一直提心吊胆，甚至可能让全场的观众陷入一种令人激动的极其的紧张之中。如果这是一场争冠的篮球比赛，那么双方都有可能获胜，取决于发生在孤注一掷地投出压哨球的那位选手的小拇指位置上的微观差异。

当我们开始讲出一个想法时，我们不知道我们最终会怎么说，也不知道我们最终会沿着哪条语法路径说下去，我们还无法预测口误或者我们在小小的口误中暴露出来的无意识。通常而言，这种暴露几乎没什么影响，但总有些时候——比如在一场工作面试中——它们能引发巨大的反响。想想当一位政客在无意识的作用下选择了一个充满政治寓意的词汇时（例如，“反恐圣战”），人们会怎样尖锐地攻击他吧。

当我们沿着一个斜坡向下滑雪时，我们不知道在下一个拐弯处会不会摔倒。每个拐弯处都有风险——有的小一点，有的大一点。一起事故有可能造成一次骨折，而我们对引发这起事故的原因根本无从推测，因为这个原因深深地埋藏在雪与我的滑雪板之间细节性的交互作用之中。

关乎我们摔倒方式的最微小的细节，都可能带来完全不同的后果，决定着我们将遭遇一次改变人生的多发性骨折，还只不过是一次微不足道的小骨裂。

简而言之，人类体验到的宏观世界是一种均质的单相混合物，其范围从最可预知的事件一直变化到极端不可预见之事。我们在生命的最初几年就熟悉了这段波谱，而大多数人类行为类型的可预知性程度，已经成为我们的第二天性。待我们走出童年之时，我们对于不可预见性在日常世界中的大多数存在之所，已经习得了一种反射水平的直觉；而这段波谱上更不可预见的一端，既在召唤我们，也让我们心生畏惧。我们被冒险所吸引，但也害怕冒险。这正是生命的本性。

闹壳㊟

下面，我要转入一个多少有些更为复杂的隐喻，借以思考在我们的大脑和思维中（如果你肯纵容我使用灵魂这个术语的话，那么最终便是在我们的灵魂中）因果性的多重层级。想象一个没有摩擦力的台球桌，桌面上有不止16个台球，还摆放着数不清的极小的弹子，我们将它们命名为“小交弹”（sim，这是“交互作用的小弹子”的缩写）。这些小交弹彼此相互碰撞，还在四壁间来回反弹。它们在一个绝对扁平的世界中相当疯狂地乱冲乱撞。而且，既然摩擦力不复存在，它们便只能不断地冲来撞去，永无停歇。

截至目前，我们的设定听起来很像是一种二维的理想气体，但我们接下来还要再加一点额外的复杂性。这些小交弹还是具有磁性的（所以让我们把它们的名字改成“小磁交弹”，simm，多加一个“磁”字，代表磁性）。当它们以较低的速率撞到彼此时，它们可以黏在一起，形成聚团；我希望你能允许我将其命名为“小磁交弹球”（simmball）。一个小磁交弹球由数目巨大（我不关心这个数字是1 000还是100万）的小磁交

弹构成，并且在边缘处频繁地损失一些小磁交弹，同时收获另外的一些。因此，这个系统里存在着两种极端不同的居民种类：微小、轻盈、飞速移动的小磁交弹，和庞大、沉重、近乎不动的小磁交弹球。

由此可知，在这张台球桌——此后将其称为“闹壳”（careenium）[注]——上发生的动态就包括了小磁交弹彼此之间的碰撞以及它们冲入小磁交弹球的撞击。当然，此中的物理细节同在标准气体中的一样，也涉及了动量传递、角动量、动能和转动能，但是对于这些方面，我们甚至连想都不会去想，因为这只是一个思考实验（“思考”的双重意义都包含在此）。对于我们的目的而言，唯一重要的是，这些碰撞一直都在发生。

小磁交弹球主义

为什么要土里土气地造出一个关于“符号”（symbol）的双关？因为我还要在我们的系统里多加入一点复杂性。构成了这一系统边界的垂直墙壁，对于外部事件的反应非常敏感（比如，有人碰到了桌子的外沿，或者甚至只是轻轻吹了一吹），会瞬时向内收缩一点。这种收缩的本性中保留了某些外部因果事件的痕迹，自然影响着自内部撞击这一部分墙壁的小磁交弹的运动方式，而且还会对邻近的小磁交弹球的缓慢运动造成间接的影响，由此，这起事件得以内化入这些小磁交弹球之中。我可以设定，一个特定的小磁交弹球总是以某种标准的方式轻轻地吹气，而另外一个则会尖锐地鸣响，如此等等。如果不深究细节的话，我们甚至可以设定，小磁交弹球的构造反映着自外部世界来犯的各种事件的历史。简言之，对于观察这些小磁交弹球并懂得如何解读其结构的人来说，这些小磁交弹球就解码事件而言，具有了符号化的意义。这就是为什么我要土里土气地造出那个双关的文字游戏的原因。

这个意象当然有点不着边际，但是我们要记住，闹壳仅仅是为理解

我们的大脑而提出的一个有用的隐喻，而事实上，我们的大脑也正是相当不着边际的。它们同样也包括了微小的事件（神经激活）和较大的事件（神经激活的模式），而后者照理应以某种方式具有表征的素质，从而令我们得以注意并记住发生在我们颅外的事情。如果你认真想一想，这种外部世界以符号化的模式在脑中的内化本是一种十分不着边际的观念，可我们知道，在进化的压力之下，它确实以某种方式成为一种存在。如果你愿意的话，大可以把那个闹壳也想象成同样是经由进化而成。你可以想一想，几十亿个更原始的系统在世界中为了生存而竞争，而这些闹壳的出现，就是这番竞争的结果。但是，我们在此不需要关心这些闹壳的进化本原。关键的要点是，虽然没有任何一个小磁交弹能够独自编码任何信息或者发挥任何一种符号化的作用，小磁交弹球却能在它们所处的远为宏观的层面上，真正地实现编码和符号化。

以还原主义视角看待闹壳

一名现代物理学家在听到这个故事的时候，首先会倾向采用一种还原主义，把大的小磁交弹球贬损为仅仅是一种副现象（epiphenomena）。意思就是说，虽然不可否认它们在那里，但它们对于理解那个系统却并非关键，因为它们是由小磁交弹组成的。在那个闹壳中发生的每一件事都只能以小磁交弹为单位来进行解释。而这毫无疑问是正确的。一座火山也不可否认地在那里，但是有谁需要去讨论关于山脉、地下压力、喷发和岩浆这类事物呢？我们可以把这些副现象性的概念一股脑儿地丢掉，转换到原子或基本粒子所在的更深层级。至少对于我们的物理学家来说，底线在于，副现象只是便利式的速记法，概括着大量更深和更低等级的现象；它们从来都不是任何一种解释的枢要所在。还原主义在此！

当我们抛弃一切宏观的说法与看待事物的方式时，唯一的问题就是，复杂度也会随之急剧增高。如果我们拒绝使用任何涉及副现象的语

言表达，那么我们就注定只能看到无数的粒子，而这显然并不是一种受人欢迎的想法。

另外，当人们眼中只有数不清的粒子时，在这个世界里就不再有自然的清晰边界了。人们不能在火山周围画一个圆圈，然后宣称："只有在这片地界内的粒子才被包括在内，"因为粒子不会尊重任何这种宏观的线条，就好比蚂蚁也不会尊重人类仔细测量并精确划定的地界线。宇宙中没有任何一个固定的部分可以严格地与其余部分隔离开来，而不产生任何交互作用——甚至连大约如此也做不到。在还原主义者看来，以不可侵犯的宏观时空界线，把宇宙切分成不同的地带，是一种完全不可理喻的观念。

下面这个例子，突出地表现了局部时空界线的无意义性。在1993年11月，我读到了几篇关于一颗向木星"缓慢"移动的彗星的新闻报道。虽然距离初始时刻还有大概八个月之久，天文物理学家已经预测了它撞上木星的具体时间和地点，时间上即便没有精确到秒，也至少精确到了分钟。这个有关距离地球几十亿英里远而且不可及的某颗彗星的事实，已经在我们星球的表面上产生了巨大的影响，成群结队的科学家投入计算它抵达木星的时间，报纸和杂志把它印上了封面故事，还有数以百万计的和我一样的人们正在阅读这些故事。在这些人之中，有一些可能因为对故事太过着迷而错过了自己的航班，或者可能因为对此共同的兴趣而结识了新的朋友，又或者也可能因为重读报刊文章中的一个短语而晚一秒到达了下一个红绿灯，不一而足。当初始时刻终于到来，彗星完全如预测一样地撞到了木星的远侧时，地球上的居民对这起遥远的宇宙事件投入了巨大的关注。可想而知，在彗星撞上木星的数月之前，在我们的星球上就已经发生了某些如果没有彗星的到来则不会发生的意外事故，还出现了某些为此而意外怀上的宝宝，某些苍蝇因此被拍死了，某些咖啡杯因此被打碎了，不一而足。所有这些发生在我们这颗小星球上的疯狂的事，全是由一颗彗星造成的，它从几十亿英里开外的寂静空间穿行滑过，距离与一颗巨大行星的遭遇还有接近50万分钟的时间。

我想以此说明的要点在于，如果走上一条纯粹的还原主义路线，就会陷入很大的麻烦之中；不仅“系统”中所有的对象都是微观的，而且多到难以计数，而且这种系统本身的生长也无视时间与空间的界限，并最终成为永远都要纳入考虑的整个宇宙。既然所有一切都分裂成了数以亿万计的目不可见的小碎片，散落四方，可理解性便荡然无存了。还原主义是残酷无情的。

站在较高层级的视角上看待闹壳

换言之，如果在副现象的层级上，存在一种可供感知与理解的“逻辑”，那么我们人类定会争先恐后地跃入那个层级。事实上，我们根本没有选择。所以，我们确实就在谈论火山、喷发和岩浆这类事物。同理，我们会谈论咬手指的习惯、黑麦面包、苦笑以及犹太人的幽默感，而不会谈论细胞和蛋白质，更不会谈论原子和光子了。毕竟，我们本身也是相当巨大的副现象，而正如我在本书中多次提出的那样，这个事实已经判定我们必须通过与我们的尺寸差不多级别的其他副现象（例如，我们的父母，我们的猫、车和蛋糕，我们的帆船、萨克斯管和檫树）来谈论世界。

下面，让我们返回闹壳，来谈谈那里究竟发生了些什么。到目前为止，我对于它的描绘还都聚焦在那些小磁交弹和它们的冲击相撞之上。小磁交弹球也出现了，但是它们的功能与墙壁类似，巨大而静止不动，供小磁交弹的碰撞弹击。在我的心目中，我经常把那些小磁交弹的动作想象成跟一台弹球机里的银色弹球一样，而小磁交弹球则相当于“弹钉”，也就是弹球沿着游戏斜板滚下来的过程中遇到的静止不动的更大的圆柱体，弹球会在上面发生撞击和反弹。

但是接下来，我要介绍一种看待闹壳的不同方式，其以两种感知转移为特点。第一，我们要转移到延时摄影的成像之上，这就意味着，那

些几乎无法察觉的慢动作会加快速度，而进入可以察觉的视线之内，而快速的动作会变得更快，以至于它们看起来甚至都不是一片模糊的幻影，而是完全无法察觉，就跟一台电风扇转动的扇片一样。第二种转移是，我们从空间上向后退，或者说变焦缩小成像。如此一来，小磁交弹就会变小到肉眼难以察觉，而小磁交弹球就必然成为我们注意力的唯一焦点所在。

现在，我们在那张桌面上就看到了一种完全不同的动态类型。我们看到的不再是小磁交弹撞向看起来纹丝不动的巨大聚团；恰恰相反，我们意识到那些聚团根本就不是静止的，而是拥有属于自己的生命。它们在桌面上来来回回地移动，彼此交互作用，仿佛桌面上除了它们之外没有任何其他的东西。当然，我们心里明白，这一切的发生都多亏了极其微小的小磁交弹于深底处的左冲右撞，但我们已经不再能看到那些小磁交弹了。在我们看待事物的新方式中，它们在桌面上癫狂的冲撞仅仅构成了一个静止的灰色的背景。

想想桌上一个玻璃杯里盛的水何以在我们的眼中是完全静止不动的。而如果我们的目光可以转移层级（可以想想控制望远镜内外伸缩的转环）而让我们得以在微观的层面上端详那杯水，我们就会意识到，它根本一点儿都不平静，而是一团水分子来回冲撞的疯狂喧嚣。其实，只要在水中加入一些胶质粒子，它就成为一个布朗运动的场地，胶质粒子一刻不停地随机搅动，反映的正是其与远远比其更加微小的水分子之间发生的不计其数而又难以察觉的碰撞。（胶质分子在此扮演了小磁交弹球的角色，而水分子则发挥了小磁交弹的功能）。布朗运动这种在显微镜下可以观察到的现象，由爱因斯坦在1905年利用分子理论进行了极为详尽的解释[注]。当时，分子还只是一种存在于理论假设中的实体，可是爱因斯坦此番解释的影响太过深远（而且，更重要的是，它与实验数据吻合一致），竟成为对分子真实存在性的最为重要的确证之一。

在闹壳中，是谁在使唤谁？

如此，我们终于抵达了问题的枢纽所在：这两种看待闹壳的方式，哪一种才是正确的？或者可以呼应罗杰·斯佩里的那个关键问题发问，在占据着闹壳的因果力量种群中，是谁在使唤谁？根据其中一种观点，没有意义的小磁交弹是主要的实体，如癫狂一般呼啸飞驰，缓缓地推动着那些沉重而消极的小磁交弹球来回移动。在这种观点下，是那些微小的小磁交弹在使唤着大的小磁交弹球。这就是它的全部了。事实上，这个观点甚至根本就没有把小磁交弹球认可为一种分别的实体，因为我们有关它们动作的所有说法，都只是在用一种梗概式速记法在谈论小磁交弹的所作所为。从这个视角出发，这里没有小磁交弹球，没有符号[注]，没有观念，也没有正在发生的想法，而只有大量微小而闪光的磁性球体在一团混沌与喧嚣中没头没脑地四处乱撞。

根据另外一种观点，当经过了加速与伸远之后，所有那些闪亮而微小的小磁交弹，都融为一团分辨不清的灰汤，而表现出丰富的交互作用的小磁交弹球则吸引了全部的兴趣。人们看到成群的小磁交弹球以一种与围绕着它们翻滚的灰汤完全无关的“逻辑”触发着其他的小磁交弹球；当然，除非从一种相当乏味的角度出发，可以说小磁交弹球的能量取自那无处不在的汤中。确实，并不足以令人惊奇的是，小磁交弹球的逻辑只与小磁交弹球所象征的概念有关。

小磁交弹球的舞蹈

从我们徘徊在桌面之上时所采取的高层级的宏观有利视角出发，我们可以看到一些观念引发了另一些观念，我们可以看到一件符号化的事件提醒另一件符号化事件的体系，我们可以看到小磁交弹球的精细模式聚合在一起而形成甚至更大的构成类比的模式。简言之，我们可以直

观地在小磁交弹球的模式化舞蹈之中发现思维的发生逻辑。也正是在这后一种观念中，在专属于它们自己的符号化层面之上，小磁交弹球在互相使唤。

小磁交弹当然还在那里，但它们仅仅是在为小磁交弹球的舞蹈做嫁衣，保证其可以发生，同时却不再与正在发生的认知过程相关。这就好比空气分子之间相互冲撞的微观细节与一架风车的叶片的转动无关一样。随便什么样的空气分子冲撞都可以——风车叶片得益于其空气动力学本性，不管怎样都会转动。同理，随便什么样的小磁交弹冲撞也都可以——“思想风车”得益于其小磁交弹球的符号化本性，不论如何都会搅动起来。

如果以上有任何一种说法让你觉得太过不着边际而无法理解，那么就请反过头来想想人类的大脑吧，想一想，为了让我们思考的逻辑得以发生，在大脑内部一定要发生哪些事情。除了像我们在本章中所讲的故事一样的事情之外，在每一个人类的头颅之内，还发生着哪些其他的事情？

显然，我们又回到了很久以前被放回书架的那本书的标题让我产生的疑问，也是罗杰·斯佩里曾经问过的问题：在这里面，是谁在使唤谁？答案是，这完全取决于你选择聚焦于何种层级之上。正如在一种层级上，可以合乎情理地说，是641的质数性在多米诺链的网络中使唤着多米诺骨牌，这里也同样存在着一种层级：在此之上，可以合乎情理地说，是附着于不同小磁交弹球之上的意义在使唤着小磁交弹球。即便这一切看上去颠三倒四——其必然如此，可是它仍旧与物理法则的基本因果性保持着完全的吻合一致。

---

1. 我选择这个古怪的整数641，是因为它在数学的历史中扮演了一个重要的角色。费马猜想所有可以写成$2^{2^n}+1$形式的整数都是质数，但是欧拉发现641（本身是质数）能够整除$2^{2^5}+1$，由此驳倒了费马的猜想。

2. See [Pattee], [Holland 1995], [Holland 1997], [Andersen], [Simon], and Chapter 26 of [Hofstadter 1985].

3. 道吉（Douggie），指侯世达。——编者注

4. Chapter 25 of [Hofstadter 1985] is a lengthy Achilles-Tortoise dialogue spelling out the careenium metaphor in detail.

5. “闹壳”在这里既表意，又与英文的“knock”（冲撞）谐音。在这个隐喻中，作者生造了一系列新词，每个词都是由不同英语单词的首字母组合而成的，如sim，simm和simmball，这些词由小写字母构成，所以不是缩写，而是全新的单词。为了照顾对英语不敏感的读者，这些新词的对应翻译也采用汉语单词首字组合的形式，每个字都提示着这个新词的含义，从而避免混淆不清。特别需要指出的是，小磁交弹球（simmball）与英文中的符号（symbol）谐音。这是作者颇为得意的一个小机关，可惜在中文译本中难以再现。——译者注

6. See [Hoffmann] and [Pais 1986].

7. This view approaches the extreme reductionist philosophy expressed in [Unger 1979] and also in [Unger 1979].

第4章

环路、目标与环路中的漏洞

冲水马桶的欲望

当第一批内载反馈机制的机械系统被设计出来的时候，一整套全新的观念开始成为人性关注的焦点。詹姆斯·瓦特的蒸汽机调节器是最早出现的此类系统之一；继而又出现了多得数不清的此类系统，包括调节冲水马桶上水的浮球机制、跟踪导弹的内部技术以及恒温调节器。既然冲水马桶可能是我们最为熟悉、也最容易理解的一个，我们就先来聊聊它。

抽水马桶里有一根往水箱里注水的管子，随着水位的升高，一个中空的浮球会跟着上升。这个上升的浮球连在一根硬棒上，硬棒的另一端是固定不动的，所以它倾斜的角度就反映了水箱里的水量。这个可变的角度控制着一个阀门，调节着管子中的水流。因而，当注水到达临界水平线时，这个角度会抵达一个临界值，此时阀门就会彻底关闭，由此也关停了管子中所有的水流。然而，如果水箱漏水，水平线逐渐下降，浮球自然也会随之下落，那么阀门就会打开，而向内灌注的水流也会重启。这样一来，我们有时候便陷入一种循环的境地：冲水后，如果小橡胶组件没有严丝合缝地落在水箱下水口的正中央，那么水箱就会缓慢地漏水几分钟，之后又会忽然重新开始注水，而过了几秒钟之后，注水停止，水箱又开始慢慢漏水，几分钟之后，还会再注水几秒钟，这就与呼吸的循环模式多少有点儿相似，循环往复，永不停止——直到有人轻轻摇晃马桶的手柄，希望连带着摇动橡胶组件，让它落回下水口上合适的

位置，从而解决漏水问题。

有一次，我在离家度假的几周时间里请我的一位朋友帮忙照看房子，他在第一天冲了一次厕所，而那个小橡胶组件恰好没有落到中心，因此便触发了这种循环。我的朋友兢兢业业地数次返回，检查房子是否安然无恙，可他从没有注意到有任何异常之处，所以在我整个离家的期间，那个马桶水箱一直在漏水和重新注水之间循环，导致我多付了500美元的水费！难怪人们那么不信任反馈环（feedback loops）！

我们可以拟人化地把冲水马桶描述为一个“尝试”让水面达到并维持在一条特定水平线上的系统。当然，绕过这种拟人化的语言也并非难事，因为我们不费吹灰之力就能看明白这个机制运作的原理，而且相当明晰的一点是，这样一种简单的系统是没有欲望的；即便如此，当应对一个水箱漏水的马桶时，人们还是会忍不住地说，这个马桶正在“尝试”把水位提升到标准的水平，可是它“做不到”。人们无法真正地把欲望或绝望投射于这个设备；那只是一种言说方式，也是一种颇具便利性的简略表述。

一个名为欲望的足球

采用这样一种以目标为导向——即目的论式——的简略方式(注)来描述具有反馈的系统似乎于我们很有吸引力，可为什么换作结构偏弱的系统，就没那么吸引人了呢？这完全取决于一个系统的“感知”（可以说）反馈到其行为之中的方式。当某个系统总是朝着一个特定的状态移动时，我们就把那个状态视为该系统的“目标”。正是这种系统的自我监管与自我调控的本性，使得我们启用了目的论式的用语。

但是，哪些种类的系统具有反馈、持有目标并拥有欲望呢？沿着多草的山坡向下滚动的一个足球，会“想要”到达山脚吗？我们中的大多数

人都会毫不迟疑地回答说“不会”，而在事物为何运动这一原始的亚里士多德式概念面前，不由自主地退避三舍。但是让我们把这个情境稍微改变一下，然后再问一次这个问题。

我们把足球向下滚落的场所换成一个横截面为U形的狭长的路边水渠；此时，这只足球有在追求任何目标吗？这样的一个足球，在其加速向前滚动的同时，首先将滚上水渠的一侧，然后落回中心，穿过中点，再滚上另一侧，然后又落下，再上去，如此往复，从一条围绕水渠中心来回摇摆的正弦曲线的路径，逐渐汇合于水渠底部一条笔直的路线。这里是否存在“反馈”呢？这只足球在“追寻”水渠的中心线吗？它“想要”沿着水渠的斜谷滚动吗？这个例子与前面沿着山坡滚下的足球的例子共同表明的是，反馈、目标或欲望的在场或缺席，并不是一件非黑即白之事；这些都是本能的判断。

目的论的光滑斜坡

当我们遇到反馈更加繁复而机制又更加隐秘的系统时，换用目的论用语——首先是目标的用语，然后是“愿望”、“欲望”与“尝试”——的倾向变得对于我们而言更具诱惑力，更加难以抵抗。甚至不需要反馈有多么繁复，只要其机制是隐藏的，就会如此。

在旧金山探索博物馆（Exploratorium museum）里有一块圈起来的场地，人们可以站在那里，观看一个红色亮点在墙壁和地板上来回舞蹈。如果有任何人想要伸手触摸那个红点，它都会在触碰发生前的最后一瞬间疾速躲开。事实上，它来回舞蹈的方式，仿佛是在戏弄追逐它的人——有时会完全停下来，向人们发出挑衅，怂恿他们伸手追来，然后在间不容发之际，一掠而逝。然而，不管表面看来如何，背后并没有任何操纵它的人——只有某个简单电路中的简单反馈机制，在监控圈地区域中的事物，并控制着那光束。但是，那个红点在全世界看来都似乎具

有一种人格，拥有一种戏弄别人的顽皮欲望，甚至是一种幽默感！可以说，探索博物馆中的这个红点看上去比一只苍蝇或一只蚊子更有生气，苍蝇和蚊子也会尝试避免被人拍到，但显然并没有任何足以察觉的幽默感。

在卡尔·西姆斯（Karl Sims）拍摄的名为《虚拟生物》（*Virtual Creatures*）[注]的视频中[注]，有一些由几根（虚拟）管状物铰接在一起的虚拟生物。这些生物可以“摆动”它们的肢体，因而得以在一块（虚拟）平板上移动。当赋予它们一种最基本的知觉，并设定一个驱使它们获取特定种类资源的简单反馈环时，那么它们获取资源以及与这些资源的“竞争者”展开搏斗的方式，会给观众带来一种毛骨悚然的观感，仿佛正在目睹卷入生死之战的原始的生灵。

回到我们更熟悉的一个层面上说，当有些植物——想想一朵向日葵或者生长中的葡萄藤——以正常的速度观察时，它们看起来如岩石一样静止不动，因此也明显缺乏目标；但是，当通过延时拍摄来观察时，它们似乎突然之间对周遭的环境高度敏感起来，也具备了清晰的目标和达成目标的策略。问题在于，这种没有大脑的系统，是否也灌注着目标与欲望。它们拥有希望和期盼吗？它们拥有恐惧与梦想吗？还有信念与哀伤呢？

一个反馈环的在场，哪怕其相当之简单，也会对我们人类构成一种强大的压力，迫使我们把描述的层级从机械的无目的性层级（在其中，推动事物运动的是力）切换到控制论的以目标为导向的层级[注]（在其中，直白地说，推动事物运动的是欲望）。正如我在前文曾强调过的那样，后者只不过是一种对于前者更有效率的重述；尽管如此，随着系统拥有的反馈环越来越精妙繁复，这种简略表述的有效性也会变得几乎难以抗拒。最终，不仅这种目的论式的语言变得不可或缺起来，我们还渐渐不再能够意识到，还有任何另外的视角存在。到这一步，这一视角已经被锁定在我们的世界观之中了。

反馈环与指数增长

我们所有人都最为熟悉的一种反馈环的类型，也很可能是反馈环得名由来的情况，是声频的反馈。发生这种反馈的典型场合是在一座礼堂里，当麦克风离一个正在播送声音的扩音器太近时，经过放大的声音又会被麦克风收取。某个声音输入后（任何声音都行——没有区别），更加响亮地输出，然后那个声音再次输入，又更加响亮地输出，然后又重新来过一次。而忽然之间，你几乎凭空地得到了一个恶性循环，它制造出可怕的刺耳的尖锐声音，让观众们纷纷捂住自己的耳朵。

这种现象太过司空见惯，以至于似乎不需要更多的评述，可实际上，这里面还是有两件事是值得指出的。第一，理论上，任何一种输入声音的每一次循环都会按照一个固定的因数放大它的音量。设这个固定因数为k，则两次循环的放大因数就是k^2，而三次循环则是k^3，依次类推。没错，我们都听说过关于地球人口指数增长的恐怖故事或者其他类似的灾难，从而得知了指数增长的力量。（在我的童年时代，指数的力量就通过一个不那么有害的故事，在我的心中打下了同样不可磨灭的烙印。在故事里，一位苏丹国王下令[注]在国际象棋棋盘的每个格子上放入比前一个格子里多一倍的米粒——当棋盘尚未填满一半时，局势已然明了，在苏丹的领地内，乃至在全世界的范围里，都找不到足够多的米粒，可以哪怕只是接近布满这个棋盘。）既然如此，从理论上讲，最轻微的一声耳语也会很快放大为一声咆哮，继而没有限度地不断增大，首先让礼堂观众席上的每个人失聪，接着马上开始猛烈地震动建筑的椽子，直到它在耳聋观众们的头顶倒塌下来，然后，再经过几次循环之后，星球都会被震成两半，而最终，整个宇宙都会被消灭。这幅末日景象的似是而非之处在哪里呢？

一代谬误

这幅图景中最首要的谬误是，我们没有考虑执行指数过程的实际设备，即音响系统本身，特别是那个扩音器。我要是想用一种最粗暴的方式表明我的观点，只需要提醒你，在礼堂屋顶倒塌的时刻，它会落在扩音器上，把它砸成碎片，旋即中止了那个失去控制的反馈环。那个小小的系统，自身就包含了自我毁灭的种子(注)。

但是，在这幅场景中，依然存在某些似是而非的东西，因为我们都知道，事情绝不会发展到那个地步。礼堂绝不会倒塌，而观众们也不会被巨响震聋。有某种东西早早地就拖慢了这失控的过程。那是什么东西呢？

二代谬误

在我们推理中的另一个谬误，同样包括了音响系统的一种自我毁灭类型，但是与被砸得支离破碎相比，要来得更加委婉一些：随着声音越来越响，扩音器将不再按照那个恒定的因数k进行放大。当音量到达某个水平后，扩音就开始失灵了。这就好比一辆在地面飞驰的轿车并不会持续不断地以一种恒定的加速度加速（每小时增加100英里，然后是200英里、300英里、400英里，很快就打破了音速的阈限），反而最终会在某个速度峰值处稳定下来（这是路面摩擦力、空气阻力、发动机内部限制等等因素的共同作用）；同理，一台扩音器也不会一以贯之地放大任何音量的声音，而是最终会达到饱和，放大的程度越来越低，直到抵达某种音量的水平，让输出的声音和输入的声音具有同等的音量，而一切也将稳定于此。放大因数变成1时的音量，便是我们熟悉的那种令你抓狂却并不会震聋你的尖锐声音，这比让礼堂在你的头顶上倒塌的程度可要轻得多了。

那么，为什么它总是会发出同一种刺耳的声音呢？为什么不是一种低沉的咆哮呢？为什么不是一帘瀑布、一架直升机或者一阵悠长沉重的

惊雷的声音呢？这与该系统自然共振的频率有关。这种声音的频率可以类比于操场上一架秋千大约几秒钟一次的自然振荡频率。一台扩音器的反馈环也具有一种自然的振荡频率，而且出于我们不必了解的原因，它通常都拥有一个与高频尖叫十分接近的音高。然而，这个系统并非一下子便能精准地落在其最终的音高之上。如果你能把这个过程极大地放慢，你就会听到声音是在逐渐地逼近那个尖锐的音高，很像那只滚动着寻找渠底的足球。换句话说，声音借助于在频率上一系列快速的前后摆动，仿佛近乎是“想要”抵达声谱中那个自然定点一样。

我们在这里所见的是，即使是在最简单的想象中的反馈环，也具有多层级的巧妙性与复杂性。我们平时对此几乎没有怎么思考过，谁料这里的发现如此丰富，充满了惊奇。那么想象一下，在更复杂的反馈环那里，又将发生何种的景象。

反馈及其坏名声

在20世纪70年代的某一天，我的父母第一次想要买一部摄像机，而我跟着他们来到了商店。我们想看看他们都有些什么，于是被领到了商店的一个区域，那里的架子上摆着几台电视机屏幕，而一台摄像机正插在其中一台的背面，让我们得以看到摄像机镜头捕捉到的画面，并以此评估它的色彩准确度以及诸如此类的性能。我拿起摄像机镜头，对准了我的父亲，然后我们便看到他那逗人的微笑一下子跃上了屏幕。然后我又用镜头对着自己的脸，而倏忽之间，说变就变，屏幕上父亲的脸换成了我的脸。而在这之后，我情不自禁地产生了一种冲动，想要试着把摄像机的镜头对准电视屏幕本身。

真正奇妙的事实至此方始到来，而我将永远带着某种程度的羞耻感将其铭记于心：我在闭合这个反馈环之前，竟然迟疑不决！我没有直接上手去做，反而犹豫畏缩起来，胆怯地向售货员寻求这么做的许可。我

到底为什么会做出这样的事情来呢？嗯，也许，如果我复述一下他对于我的请求是如何回复的，这件事就容易理解了。他说的是：“不，不，不行！不要那么做——你会把摄像机搞坏的！”

那么，面对他的惊惶，我是如何回应的呢？轻蔑？嘲笑？终究还是径自照自己心血来潮的念头去做了？非也。事实的真相是，我对于自己并没有那么确信，而售货员惊恐的反应，更加强了我模模糊糊的局促感，所以我抑制住了自己的欲望，并没有那么做。可是后来，在我们抱着新买的摄像机开车回家的路上，我又仔细想了想这件事，可就是无论如何也看不出我闭合了那个反馈环会给那个系统带来什么的危险——无论是摄像机还是电视屏幕（虽然从直觉出发，它们都可能面临致命的毁灭）。所以，当我们回到家后，我小心翼翼地试着把摄像机镜头对准了屏幕，妙哉怪哉，完全没有发生任何可怕的事情。

我猜，引人担忧的危险应该是某种类似声音反馈之事：也许屏幕上一个特定的点（当然，就是摄像机正对着的那一点）会变得越来越亮，越来越亮，而屏幕很快就会从那里开始融化而毁坏。但是，为什么会发生这种事呢？如果和声音的反馈一样，那么这里就必须发生某种对于光的强度的放大，可我们知道，摄像机本来也不是被设计出来放大图像的，它只是把图像转移到另一个地方而已。正如我在回家路上的冷静思考中明白的那样，标准的视频反馈根本就不存在任何危险（顺带一提，我并不知道“视频反馈”这个说法是什么时候由谁提出的；可在当时，我肯定是闻所未闻）。但是，有危险也好，没有危险也罢，我总之清晰地记住了我在商店时的犹豫不决，因此也很容易想象出那名售货员的惶恐，尽管那是毫无道理的。反馈——让一个系统回转或扭转至自身，由此形成某种神秘而忌讳的循环——这看起来是危险的，这看起来是不必要的冒险，也许甚至从本质上就是错误的，而不必管这么说到底意味着什么。

这些都是原始的、不理智的直觉，谁也不知道它们来自何方。也许

有人会揣测说，对任何一种反馈的恐惧，都只不过是一个人对于声音反馈体验的简单而自然的泛化，但是我对此抱有一种莫名的怀疑，觉得恐怕没那么简单。我们都知道，有些部落害怕镜子，很多社会群体都不信任摄像机，还有特定的宗教禁止给人们画像，不一而足。再现人的自身被视为可疑的、古怪的，也许最终还是致命的。看起来，这种对于循环的不信任感，流淌在人类的血液之中。尽管如此，同很多如悬挂式滑翔和跳降落伞等极限运动一样，我们之中有些人会被它深深地吸引住，而其他一些人仅仅是想一想，都会吓得要死。

上帝、哥德尔、变音符与迷思

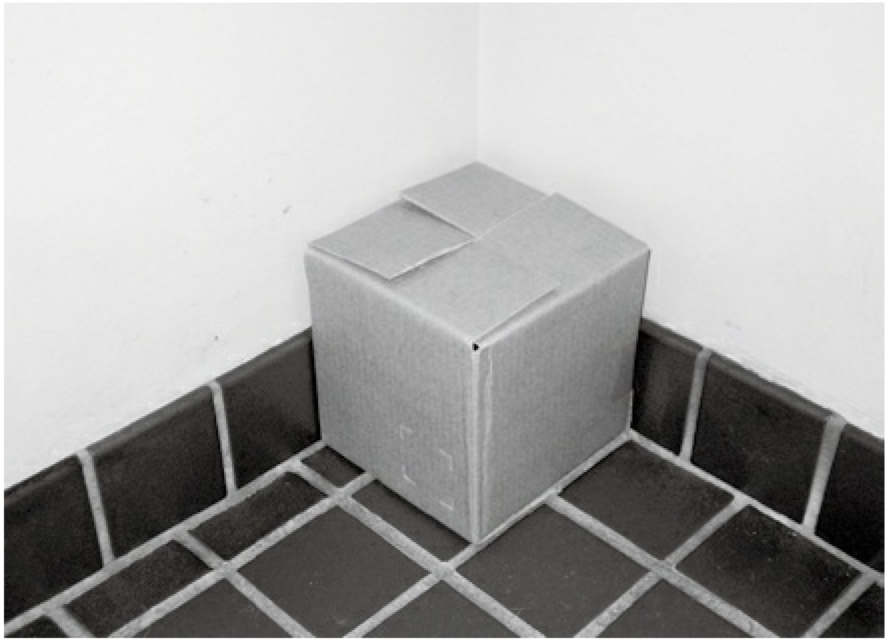
14岁的时候，我在一家书店里随意浏览，撞见了一本题目为《哥德尔证明》（*Gödel's Proof*）的小开本平装书[注]。我完全不知道这位哥德尔是何方神圣，也不知道他（我敢肯定，自己在那么小的年龄，处于人生的那个阶段时，不会想到“他或她”[注]）可能证明了什么，但是整本书只讲述一则数学证明——不管是何种数学证明——这一观念，激起了我的好奇心。我还必须承认，隐含于“Gödel”中显眼的“God”（上帝）一词以及栖息在“God”中央上方那个看起来很神秘的变音符，无疑也增添了些许兴味。我大脑中的分子已经被得当地挑逗起来，把信号向下传递给我的胳膊和手指，我顺手拾起了那本有变音符装点的书，匆匆翻动纸页，看到了像“元数学”、“元语言”和“不可判定性”这些撩人心弦的词汇。然后，我很高兴地看到，这本书讨论了悖论式的自我指涉句，比如“我在说谎”及其更加复杂的近亲变体。我可以看出，不论哥德尔证明的是什么，其所关注的焦点本质上并非数字，而是聚焦于推理本身；而最令人惊叹的是，对于数学本性的推理，使用的是数字。

虽然在一部分读者看来，下面这些话可能听起来有点不可理喻，但是我确实记得自己被一则长长的脚注特别地吸引住了。这则脚注讲解的

是以引号区分使用与提及的正确用法。这本书的作者——欧内斯特·内格尔（Ernest Nagel）和詹姆斯·纽曼（James Newman）——在脚注中以“芝加哥是一个人口众多的城市”和“芝加哥是个三音节词”这两个句子为例，称前者是对的而后者是错的。他们的解释是，如果想要讨论一个单词的属性，就必须使用它的名字，即把它放到引号之中。因此，“‘芝加哥’是三音节词”就是对的，因其涉及的并非一座城市，而是它的名字。两位作者继续讨论了在形式推理内部仔细做出这种区分的必要性，并指出名字本身也有名字（可以使用引号进行表达），如此无穷无尽。所以，这也就是一本关于语言如何能够讨论讨论其自身的语言（以此类推）以及推理如何推理自身（以此类推）的书。我的魂儿都被勾住了！虽然在那个时候，我丝毫不知哥德尔的定理是什么，但是我已经知道，我非得读这本书不可了。构成这本书的分子已经成功地设法令构成我头脑的分子指挥构成我手的分子去拿起构成我钱包的分子……可以了，你知道是怎么回事。

品味循环与自应用

通读内格尔和纽曼的那本引人入胜的小册子时，在我眼中显得最为迷人而魔幻的，是数学似乎在原路返回自身、吞食自我并在自己的内部螺旋上升的方式。我对于这类循环现象总是难以抗拒。比如，我在童年很早的时候，就迷恋上了通过把纸板箱的四片折盖以一种“循环”的方式互相折叠起来而合上箱子的观念——A在B之上，B在C之上，C在D之上，然后D又在A之上。这种对于悖论性的摄取令我心醉神迷。



我还总是喜欢站在两面镜子之间，看隐含的无穷无尽的镜像渐渐地消失于远方（照片由凯莉·古特曼拍摄）。一面镜子照出另一面镜子，还有比这更加诱人的观念吗？我也喜欢那张“莫顿盐”（Morton Salt）女孩抱着一个“莫顿盐”盒子的图片，她手中的盒子上画着她自己，也抱着盒子，如此不断往复，图案越变越小，却隐含着无止无终。

多年后，我带我的孩子们去好莱坞参观一个名为“马德罗丹小人国”（Madurodam）(注)的公园，里面有数十座全荷兰知名建筑的精致优美的迷你复制模型。但最让我失望的是，我没有看到马德罗丹小人国自身的迷你复制品，其中自然也应包括一个更小的复制品，如此无穷无尽……令我尤为惊讶的是，在全世界那么多地方，这个缺憾竟然发生于荷兰——那不仅是艾舍尔（Maurits Cornelis Escher）(注)的故乡，而且是德罗斯特（Droste）的著名热巧克力的家乡。那种巧克力的盒子和“莫顿盐”的盒子极像，也以一种无限回归的方式包含着自身，而且在荷兰家喻户晓。



我对这种循环的痴迷，其根源可以追溯到很久以前。当我还只不过是个小屁孩儿的时候，大概是四五岁左右吧，我就自己琢磨出，也可能是有人告诉我，2加2等于4。那个迷人的短语——“两个2”——让我的脊背打起了阵阵激灵，因为我意识到，它涉及将“2”这个概念应用于其自身之上。这是一种自我指涉的运算（operation），把一个概念扭转回到它的自身。我就像一个胆大包天的飞行员或攀岩者一样，渴望着更多类似的、以及更具风险的体验，于是我自然而然地对自己发问，三个**3**会得出什么。因为年纪太小，我没法凭一己之力解开这个谜题（例如，画一个每列三个点，一共三列的方格），于是我向我的母亲，那位智慧之泉，求助答案。她平静地告诉我，答案是9。

刚拿到答案的时候，我是很开心的，可没过多久，我就模模糊糊地开始担心，自己并没有问到她正确的问题。我的困扰在于，无论是我的

新短语还是那个旧的短语，都只包含了两次对题中数字的复制，而我的目标则是要超越这种二重性。于是我得寸进尺地发明了一个更接近三重性的短语注：“三个三个3”；遗憾的是，我并不知道它是什么意思。所以，我自然而然地又一次转向我的智慧之泉去求助了。我记得，我们就这件事（幼年的我，深信这件事已超出了地球上任何人的理解）展开了一场谈话；我还记得，她向我保证，她完全理解我的观念；而且，她甚至把答案也告诉了我，可是我忘记了——当然，要么是9，要么是27。

但是那个答案并不是要点。要点是，在我最早的记忆里，就存在一种对环路结构、自应用运算、循环性、悖论式行为和隐含无穷性的热切向往。其对我而言，曾是向我滚滚而来的珍爱至宝。

懦弱的类型论

前面讲到的小插曲，揭示了我与很多人共同享有的一种人格特质，但绝非每个人都共享这种特质。我第一次遭遇这种在人们直觉上的分歧，是在阅读《数学原理》中有关罗素发明的所谓“类型论”（theory of types）的时候。《数学原理》这本名作是罗素与他的老师怀特海（Whitehead）合写的皇皇巨著，出版于1910—1913年。

早年间，罗素曾不遗余力地把数学搭建在集合论的基础之上。他深信，集合论构成了人类思想最深层的根基。但是正当他以为离自己的目标已经近在咫尺之际，他却出乎意料地在集合论中发现了一个可怕的漏洞。这个漏洞（这个词用在此正好合适）的基础是“不包含自身的所有集合的集合”这一概念，它在集合论中是合法的概念，可实际上却是高度自相矛盾的。为了把这一发现的致命性传达给更广泛的受众，罗素生动而形象地把它类比为一个假想中“只给本村中不给自己刮脸的人刮脸”的理发师。设想这样一位理发师的存在是自相矛盾的，而且原因与罗素悖论一致。

发现集合论中竟容纳着此类自相矛盾的问题之后，罗素为数学寻找坚实根基的美梦就此破碎。这次创伤在他的心里灌注了一种恐惧，因为他把自己刚刚经历的这场智识灾难仅仅归咎于循环，从此便怕了那些容许自我包含或自我指涉的理论。

然后，为了努力复原，罗素偕同他从前的导师亦即眼下的同事怀特海一起，发明了一种新的集合论。在该理论中，一个集合的定义绝不会调用那个集合自身；另外，他们还设置了一个严格的语言等级，严防任何语句指涉其自身。在《数学原理》中，不会再有集合扭转回它们自身，也不再会有语言沿原路返转而归。如果某种形式语言有一个像“单词”这样的单词，那个单词也不可以指涉或应用于自身，而只能作用在处于自身层级<u>之下</u>的实体。

我在阅读这种“类型论”时，深深震惊于其远悖常识的病态式疏离(注)，及其与循环的魅力之间所保持的距离。“单词”这个单词作为“单词”范畴中的一员，到底有什么错？诸如“我在意大利多洛米蒂山脉中一个风光秀丽的村庄里开始了这本书的写作”、“本章中的正文字体是宋体”或者“这只盒子是由可回收再利用的纸板箱制作的”这些无伤大雅的句子，又能有什么错误呢？这些表达会让任何人或任何事陷于危险的境地之中吗？我看不出它们如何能够做到。

那么，“这句话包含了9个字”或者“本句最后一个字是一个笔画数为7的名词”这两句话呢？它们都很容易理解，也显然是真的，而且它们当然也不是自相矛盾的。甚至一些很傻的句子，比如“本句第9个字有10画”或“本句第10个字有9画”，也不比“2加2等于5”这句话问题更大。这三句话都是假的，或者最差也不过是无意义的表达（第二句话指涉了某个不存在之物），但是它们都没有任何自相矛盾之处。把所有的循环指涉一棍子打死，在我看来实在是一种过分偏执的招数，这也让我这一辈子都对罗素那“一朝被蛇咬，十年怕井绳”的思维深感失望。

畏惧反馈环的知识分子

多年以后，当我在为《科学美国人》（*Scientific American*）杂志撰写一个名为“元魔法主题”（Metamagical Themas）的每月专栏时，我把其中的几篇献给了语言中自我指涉的话题[注]，并于其中特别展示了由我自己、我的几位朋友以及不少读者创造出来的句子，而这丰富的句库中就包括了一些引人注目且充满挑衅意味的异想天开，诸如下列：

如果把“真”和“假”的意义互换，则此句不会是假的。

我要跟你同在两个层级之上。

下面这个句子与这句话一模一样，除了“下面”这个词与“上面”这个词交换了位置，还有“除了”与“在于”这两个词之间以及“一模一样”与“完全不同”这两个短语之间的互换。

上面这个句子与这句话完全不同，在于“上面”这个词与“下面”这个词交换了位置，还有“在于”与“除了”这两个词之间以及“完全不同”与“一模一样”这两个短语之间的互换。

这个类比就像用你自己的鞋带把你自己提起来一样。

这勾话并不旻自我指涉的，因为“这勾”不旻一个词。

如果心愿是马，那么这个条件句中前面的那个分句就是真的。

这句话每三，但仍理解。

如果你认为这句话十分费解，那么就换掉一头猪。

这个名词短语所指的东西怎么会和这个名词短语不一样呢？

文口舌殳月扁方而下口舌殳月音自。

辶勹讠氵ナイ 立而上勹讠氵ナ阝 丷。

这个短句总计[注]包括4个阴平，5个阳平，4个上声和12个去声。

虽然我从读者那里收获了大量积极的反馈（请原谅我使用这个术语），但也收到了一些极其负面的反馈。某些读者认为这过于轻浮，而这本该是一家体面的杂志才对。其中一位最为激烈的反对者，是特拉华大学的一名教育学教授。他征引了行为主义心理学家斯金纳有关自指句这一话题的论述：

> 以这种方式摆弄句子，或者分析令一句话能或不能为普通读者所接受的各种转换形式，也许没有什么危害[注]，但仍然是一种时间上的浪费，特别是当这些生造出来的句子不可能由言语行为发出的时候，尤为如此。悖论就是一个典型的例子，比如“本句话为假”，看上去若其为真则假，若其为假则真。但重点在于，没有人会把这句话作为言语行为而讲出来。在一个人说出“本句话是假的”之前，这句话必须首先存在，只有响应本身是无效的，因为直到这句话被说出来之前，它都是不存在的。

人们总归有可能有意义地说出某些自我指涉的句子，而连这种*可能性*也不假思索地否认，这样如膝跳反射式的反对，我还是头一遭遇见，也被打了个措手不及。我对于这位教育学教授的悲叹展开了长久而深刻的反思。在接下来一期的杂志上，我发表了一篇很长的回复[注]，列举了在人类的平常交流以及幽默、艺术、文学、精神分析、数学、计算机科学等领域中诸多自我指涉的案例，这些自我指涉“臭名昭著”，而又通常十分有用，甚至于不可或缺。我不清楚这位教授以及其他的对自我指涉的反对者是如何看待这些案例的。不论如何，我由此领悟了一个事实，即某些受教育程度很高并且在其他方面相当明理之人，对于自我指涉的观念或者反叠于自身之上的结构或系统，都有一种不可理喻的过敏。

我怀疑，这些人的过敏，归根结底根源于一种对于悖论或（隐喻意义上的）宇宙爆炸根深蒂固的恐惧感，有点像那个电视售货员在我威胁要用摄像机镜头指向电视屏幕时发作的惊恐。这种反馈环是我的一生挚

爱，而另一边，罗素、斯金纳、那位教育学教授和电视机售货员之流却对其敬而远之，这中间的鲜明对比给我上了受益终身的一课，即教会了我一种“类型论”——换言之，这个世界上确实存在“两种类型”的人。



---

1. See [Monod], [Cordeschi], [Haugeland 1981], and [Dupuy 2000].
2. 全名为《虚拟生物的进化》。——译者注
3. This is found easily on the Web.
4. See [Dupuy 2000], [Monod], [Cordeschi], [Simon], [Andersen], and Chapter 11 in [Hofstadter and Dennett], which discusses a trio of related “isms”——holism, goalism, and soulism.
5. Found in the charming old book [Gamow].
6. Compare this scenario of self-breaking to the story recounted in the dialogue “Contracrostipunctus” in [Hofstadter 1979].
7. Of course this was [Nagel and Newman].

8. See Chapters 7 and 8 of [Hofstadter 1985].

9. 顺便一提，此处的双引号，证明了内格尔和纽曼对于使用与提及之分重要性的坚持对我产生的毕生影响。

10. 荷兰是画家艾舍尔的故乡。艾舍尔最知名的作品是带有数学性质的版画，往往通过无限的循环和反复呈现出不可能的图案，制造出矛盾的空间。艾舍尔是本书作者另一本著作《集异璧》一书中标题中的人物，曾因该书而为更多世人所知。德罗斯特是荷兰著名的巧克力厂牌，包装盒上是一位护士端着一个放有杯子和纸盒的托盘，而杯子和纸盒上的图案跟整个包装盒的图案相同，如此无限循环。后来，人们用“德罗斯特效应”来代表一种递归的视觉形式。——译者注

11. 我当时不知道这是什么意思，不过隐约觉察到了算术运算层级的无限性，后来我才知道这叫作“阿克曼函数”。

12. 我实在忍不住要指出，《数学原理》开篇就迎来了自我指涉的大繁荣，这本书的第一句话就说：“以数学的方式处理数学的原理是本书的主题，它生发于两个相当晚近的不同学科的交叉处。”这里，《数学原理》通过“本书”这个自豪的短语指称了自己——这正是本书的作者在更正式的语境中用尽一切力量严格禁止的那类自我指涉。也许更让人感到别扭的是，禁绝自我指涉的类型理论亮相的那一章，同样以自我指涉开篇：“本章将对逻辑类型的理论进行解释，它首先是以解决特定矛盾的能力进入我们的视野……”最后还要注意的是，“我们”这个代词也是一个自指，而罗素和怀特海在使用的时候百无禁忌。他们难道对于这些反讽之处毫无察觉吗？

13. See Chapters 1-4 of [Hofstadter 1985]

14. 这种可以做到完美的自我统计或自我清点的“短句”是由李·塞洛斯（Lee Sallows）使用他自造的一台精密的模拟计算机创造出来的。我经常琢磨类似塞洛斯的短句一样的句子群，句群中的每个这样的短句都不只是在清点它们自己（比如文中那个短句对声调的统计），同时还清点某些或全部其他句子。这样一来，每个句子都要比塞洛斯的短句长出很多。不论如何，在我的幻想中，这些“个体”跟塞洛斯的经典短句有所不同，它们无法全部给出精确的报告。它们有些表达是大错特错的。在自我清点方面，我把它们想象成是相当准确的；而在另一方面，每个句子对其他句子的清点结果，准确度各不相同，有的很接近正确答案，有的完全不靠谱。无须赘言，这就是一个对于互相交往的人类社会的隐喻，每个人都有一幅相当精确的自我画像和一个不那么精确的对于他人的印象，后者通常是基于短暂而不准确的一瞥。两个“互相了解”的句子（即对于彼此的清点尽管不一定完全精准，却差不太多）就像两个好朋友，而两个对彼此的表征粗糙、不全面或缺席的句子，则是陌生人的隐喻。这个主题还有一个更加复杂的变调，那就是包含了随时间不断变化的塞洛斯类型句子的句子群。初始时，它们都分配有随机的数字，但是他们随后会开始更新迭代。具体而言，每个句子都会通过清点它们自身内部以及少数其他句子中包含的字母，并用新得到的数字替换掉原来的数据。当然，因为一切都在变动之中，所以字母的点数也不会是正确的，但是通过长时间的一系列反复重点，至少平均而言，每个句子还是有希望接近更高的准确度，而且是涉及自身的准确度，同时它们还形成了

一个小“朋友”圈子（它们相对完整和准确清点过的句子们），并与句子群中的大多数成员保持距离。这生动地刻画了我在第15章到第18章中提出的那个人们“生活在彼此之中”的想法。

15. Quoted from [Skinner] in George Brabner’s letter.

16. This is found in Chapter 1 of [Hofstadter 1985].

第5章

视频反馈

30年间，两次视频远航

20世纪70年代中期，我第一次用家里新买的摄像机镜头进行探索时，就发现了视频反馈环的丰富性。几个月后，我对于该现象的领略大为加深，于是决定对其展开一番详尽的探究，以此作为我那本《集异璧》中的一则可视化研究。我预约了斯坦福大学的电视演播室，而当我到达的时候，发现那里待人相当友好的研究员已经设置好了一台电视机和三脚架上的摄像镜头，就等我来摆弄了。把摄像头指向屏幕，实在是小菜一碟，缩进、拉远、倾斜镜头、变换角度、调节亮度和对比度，如此等等，不在话下。他告诉我，我可以随意使用这个系统，所以我就在那个下午花了几个小时，在由视频反馈环打开的“禁忌”可能性的海洋里巡游。同很多充满好奇的游客一样，我也在这场异域旅行中拍下了好多（还只是黑白的）照片。后来，我选用了其中我最喜欢的12张，用在了《集异璧》的一段对话中。

自那次在视频反馈环中的初次探险以来，已经过去了30年，科技也有所进步，所以我决定，为了自己的新书，再试一回。这一次，我受到了比尔·弗卢赫特（Bill Flucht）的鼓励与协助。比尔是基本图书公司的编辑，因为（或者说尽管）与我合作了10多年，已经成了我的一位好友。他为了这个目的，专程从纽约乘飞机过来找我。我和比尔在两家孩子以前的“游戏室”里，共同度过了几个小时的愉快时光。我们航行在那片熟悉的海域里，涛声依旧，只是所乘的船只多少更新了一点。我们最

终收获了几百张彩色快照，品质超凡地记录下了我们的航行。除了封面插图外，还有我最喜欢的16张被收入本书多处的彩色插页。

虽然两次视频航行同样斑驳多变而栩栩如生，我还是决定在本章中为很多年前在斯坦福更早的那次书写一份“日志”，因为那是我第一次探索这个现象，并逐步深入地了解了它。因此，下面的故事所涉及的电视机和电视摄像头都与制作本书彩色插页时所使用的不同，而大体上说，所使用的技术也更加古老。尽管如此，正如你将看到的那样，旧的航海日志中的大部分，依然适用于新的航行，至于少许微小的出入，我会随见随说。

一段视频旅行的日志

在为我准备的电视设备的右侧刚好有一条银色金属条带顺沿而下，而这种偶然带来了一种幸运的效果，让屏幕中套屏幕的不同层次，很容易就能辨别开来。随后，我便有了第一个发现，即存在一个临界的角度，决定着嵌套屏幕的回归是有限还是无限。如果我把摄像头对准之处由屏幕中心移向那条金属条带，那么我眼前所见就好像一张长廊右壁的快照，呈现着间隔均匀的“廊门”（这实际上是金属条带的成像），从我所“站”之地向远处延伸。但是我不能一眼看到这条“走廊”的尽头。因此，我把在这种情况下屏幕上可见之物称为一条<u>截断式</u>走廊。

如果我慢慢地把摄像头向左摇动，移向屏幕的正中，也就必然更加深入那条可见的走廊，越来越多的廊门沿着右壁出现在视野中，越来越小，越来越远——突然之间，在一个临界的时刻，便会出现一种美妙绝伦、令人神迷的无限感，而我也可以<u>一眼</u>望穿走廊，目光追着那片敞开的空虚，朝向一个单一的汇聚点（在视点理论中，这被称为“灭点”）远远地肆意延伸开去。我将此称为一条<u>无穷无尽的</u>走廊。（注意，从本质上说，在第4章自我映射的镜子照片中也能看到类似的走廊。）

当然，我看到的无穷个廊门的印象是一个错觉，因为光速与电视屏幕的分辨率为嵌套发生的数量设定了一个上限。话虽如此，与那个平庸的截断式走廊相比，窥望那似乎如魔法般无穷无尽的走廊，远为更加诱人，也更加拨动心弦。

我的下一组实验涉及摄像头的倾斜。当我倾斜摄像头时，每个屏幕都顺从地以完全一致的角度相对于容纳它的屏幕发生倾斜，于是立即制造出一条逐渐远离的螺旋形走廊——像螺丝钻一样扭转的走廊。虽然这对于眼球是很有吸引力的，但对于思维却没有什么特别的惊喜可言。

然而这里还是出现了一个意外的惊喜：当摄像头扭转到特定的角度时，我看到的不再是一条间隔着带有廊门的螺旋形走廊，反而仿佛在观看一个平面的旋涡，很像在望远镜中看到的星系。这个旋涡的边界是光滑而连续的光弧，而不再是一组组锯齿状的直线（来自电视屏幕的四边），而这种光滑性让我大惑不解；为什么锯齿状的边角会突然之间化为优美的曲线，我为此找不到任何理由。我还注意到，在每个这种“星系”的正中心，几乎总是存在一个美丽的圆形“黑洞”。（我和比尔在最近的一次视频旅行中，没能复现这种“黑洞”现象，这让我们迷惑不解，且深感懊丧，正因如此，你在本书的插页中也不会看到任何的黑洞。）。

神秘莫测的、呈展的回荡

在这段时间中的某一点，我不小心在一瞬间把自己的手挡在了摄像头之前。当然，屏幕完全变黑了，但是当我移开手时，原先的图像并没有像我预料的一样，直接跳回到屏幕之上。没错，我在屏幕上看到了一幅不同的图案，但是这幅图案同我之前所看到的所有图案模式都不相同。它并不是静止不动的；恰恰相反，它在周期性地脉动，就像一颗心脏！它的“脉搏速率”大概是每秒一圈，而在每一次短暂的“心跳”过程

中，我眼前的形状都要经历一番巨大的变化。那么，在房间里没有任何东西是正在移动的情况下，这种神秘的周期性脉动究竟从何而来？

哎哟——抱歉！我刚刚写下的那句话明显是假的——房间里的确有东西正在移动。亲爱的读者，你知道那是什么吗？好吧，正在移动的就是图像本身。这在你听来也许是一个愚蠢的、微不足道的或者自作聪明的回答，但是既然这里的图像就是呈现其本身的图像（虽然有一点点延迟），这个答案实际上是相当切中要害的。一幅呈现某种正在变化之物的图像，其自身必然也会持续地变化！在这种情况下，运动在无休无止地引发运动，因为我所应对的是一种循环性的设置——一个环路。而引起事情发生的原初运动——那份原动力——就是我那只手的运动。而由这一运动触发的视频回荡（reverberation），如今已构成一条稳定并可自我维系的可见的记忆痕迹！

这番场景令我想起了另一种被我称为“回荡犬吠”的环路现象，有时候我们能从养了很多只狗的邻居家听闻。如果一位晨跑者路过一座房子，引得一声狗吠，那么周围的狗也会跟着吠起来，接续而来的可能是涉及10多只狗的连锁反应。很快，这场犬吠派对就获得了自己的生命，而此时，这场派对无心的推动者早就已经离开这片街区了。如果狗可以变得更像机器人一点点[注]，最终不会为一遍又一遍地重复同样的事情而感到厌倦，那么它们的回荡犬吠就会成为一条稳定并可自我维系的可听到的记忆痕迹，可以追踪至那位晨跑者在房前街道上一掠而过的时候。

我在自己的视频旅行中不期而遇的动态脉动模式，与我在此之前观察到的毫无变动的“稳态宇宙”完全不同。这种稳定的、周期性的视频回荡，是我在探索潜伏于视频反馈之中的诸般可能性时，于意外中撞见的一件怪事。

即使在今天，过了这么多年以后，这种脉动的源头仍未清晰，甚至可以说，在我眼中依然神秘莫测；正因如此，它才是一种呈展现象

（emergent phenomena），不然也可以换用第3章中讨论的说法，称其为一种副现象。通常而言，一种呈展现象总归是以某种方式，依据在一个更低和更基础的层级上严格运转的法则，相当自然而自动化地呈展出来，但是发生这种呈展的具体方式，在观察者眼中却是毫不明晰的。

我承认，没能完全想明白视频回荡背后的原理，让我感到自己有一点蠢；可是，此时此地，我已经对这种现象习以为常，而它在我看来已经“讲得通”了。也就是说，我对于如何在屏幕上引发这种现象已经拥有了一个清晰的直觉认识，而且我也知道它一旦启动，就会成为一个强韧的现象，如果我不外加干涉，它可能毫无减损地持续几个小时，甚至持续到永久。我没有费尽力气去弄清楚如何在更低的现象层级上准确地解释视频回荡，反而开始干脆把它作为一个事实接受下来，并把它置于现象本身存在的层级上加以应对。这在你听来应该再熟悉不过了，因为这正是我们应对物理世界和生物世界中几乎所有事物的方式。

向环路馈送“内容”

我在一开始就提到过，斯坦福的那个装置带来了一件幸事，就是在给我使用的电视机一侧，意外留下了一条显眼的金属条带。那个条带本来是个擅闯者，却为一圈一圈循环的图像添加了一味醒目的调料，而从这个意义上来说，它便成为视频旅行第一期的一份关键原料。

在我和比尔展开视频旅行第二期的过程中，有好几次我们都惊讶地发现，那片我们正航行其中的海域对于我们而言似乎口味有点过于清淡了，而我们都向往能有更多的动作和更多视觉上的兴奋刺激。这不免让我想起在旅行第一期中，那条擅自闯入的金属条带所发挥的至关重要的“调味”功能，于是我们欣然决定，在我们的系统中也引入某种可以发挥类似功能的东西。我在房间里四处搜罗起各种各样的物体，把它们拿到摄像机镜头上来回晃荡，至于这些图像绕着视频反馈环一圈一圈循环

时会发生什么，我心里完全没有概念。我们时不时地会取得一些不可思议的结果，（又一次）完全在预料之外。例如，当我在屏幕前摇晃一串珠子的时候，呈展（选择用这个动词不是无心之举）出来的是由布满凹痕的蓝白色团构成的形状不规则的旋涡，不免让我联想到某种异域奶酪。

当然，每一样如此擅闯的东西都打开了一个全新的可能性的宇宙，因为我们可以变换它的位置以及所有其他标准变量（缩放量、倾斜度、摄像头的方向、亮度、对比度等等）。我尝试过玻璃花瓶和激光唱片这些东西，而最终又奉献了我自己的双手。你可以从本书的彩色插页中看到，这些尝试的结果是相当奇妙的；可是，哎呀，我和比尔并没有无限的时间去探索我们已经发掘并且取样过的那些多重宇宙。我们花了大约半天的时间来探索各种各样的可能性，结果从中收获了一本藏有400张照片的记忆相簿，这就是全部的成果。如同任何一次深入美妙神奇的异域之地的远足一样，我们的旅途不得不在我们尽兴之前画上句号，但我们在依依不舍之余，对于共同经历并玩味这段旅程，还是感到非常开心。

一则数学的类比

我观察到的所有那些意料之外的现象，或许正如意料之中，都依赖于屏幕嵌套（在理论上）的无限性——也就是说，它们需要那条可见的长廊是没有尽头的，而不是截断的。事实也正是如此，因为最出乎预料的视觉现象似乎总是发生在那个中心点的邻近，也就是在无穷回归（infinite regress）汇聚成一个魔法点的地方。

我的多次探索并没有告诉我，视频反馈的结果能随便生成任意一种形状，但它们确实向我展现出，我已经进入了一个比我预想之中更加丰富的可能性宇宙。如今，这种视觉丰富性令我想起了数学家本华·曼德

博（Benoit Mandelbrot）在20世纪80年代左右发现的那个令人惊叹的视觉宇宙㊟。他发明视觉宇宙时，正在研究由$z \to z^2 + c$定义的简单循环的属性，其中c是一个固定的复数，而z则是一个复数变量，其初始值设为0。这是一个数学上的反馈环，输入z的一个值，则会有一个新值输出，而这个新值又已准备好再次馈入，正与音频反馈环或视频反馈环如出一辙。关键的问题在于：如果由你来扮演麦克风或扬声器（或者摄像机与电视机）的角色，一遍接一遍地反复执行下去，那么你所得到的z值，到底会不受约束地增长，驶向浩渺的蓝色远方（或红色远方或黄色远方）㊟，还是恰恰相反，将追踪目标定位在一个有限值之上？

我们在此并不关心细节；基本的要点是，问题的答案以一种精妙的方式取决于参数c的值。如果你根据z的散度，用颜色标示出不同的c值，那么你做出的这张图将会为你展开一幅令人惊叹的画面。（这也就是我为什么要开那个浩渺的“红色远方”和“黄色远方”玩笑的原因）。不论是在视频反馈里，还是这个数学系统中，一个非常简单的循环过程就可以生发出一簇切实无法预料的回旋模式，并精妙复杂到令人难以置信的地步。

“锁定”现象

从诸如视频反馈等循环过程中呈展出来的那些神秘而又强韧到不可思议的现象，从现在起将成为本书中最主要的隐喻之一，帮助我进一步扩展有关意识与自我这一核心问题的讨论。

我从自己的视频之旅中感受到了视频反馈现象巨大的丰富性。说得更加具体一点，我学到的道理的是，在很多情况下，存在于屏幕之上的美妙的复杂结构与模式的来源，在人类观赏者眼中完全是模糊不清的。让这些模式得以存在并维系的，是系统的循环性——环路模式的存在——这一事实令我震撼不已。每当一种模式映上屏幕，那么要想为它

居于其上进行论证，唯一需要的就是乔治·马洛里（George Mallory）那句经典的俏皮话。他在解释自己为什么非得攀登珠穆朗玛峰时说："因为它就在那儿。"当牵涉到环路模式的时候，循环论证便成为重中之重。

换句话说，反馈带来了一种新的抽象现象，可以称其为"锁定"。从一个小到不能再小的暗示（在第一秒钟最早的一小段时间里发送到电视屏幕上的第一幅图像）出发，几乎立刻（在大约20次或30次循环之后）就实现了这一暗示中所囊括的全部蕴含——而这种新的更高层级的结构、这种在屏幕上呈展的模式、这种副现象，也就在循环的作用下被"锁定"了。它不会逝去，因为它永远都在自我更新、自我馈送和自我再生。也可以说，呈展的输出模式是一种自稳定结构，尽管反馈环本身足够简单，这种结构的来源却几乎无从理解，因为它已经绕了太多圈。

视频反馈呈展的新现实

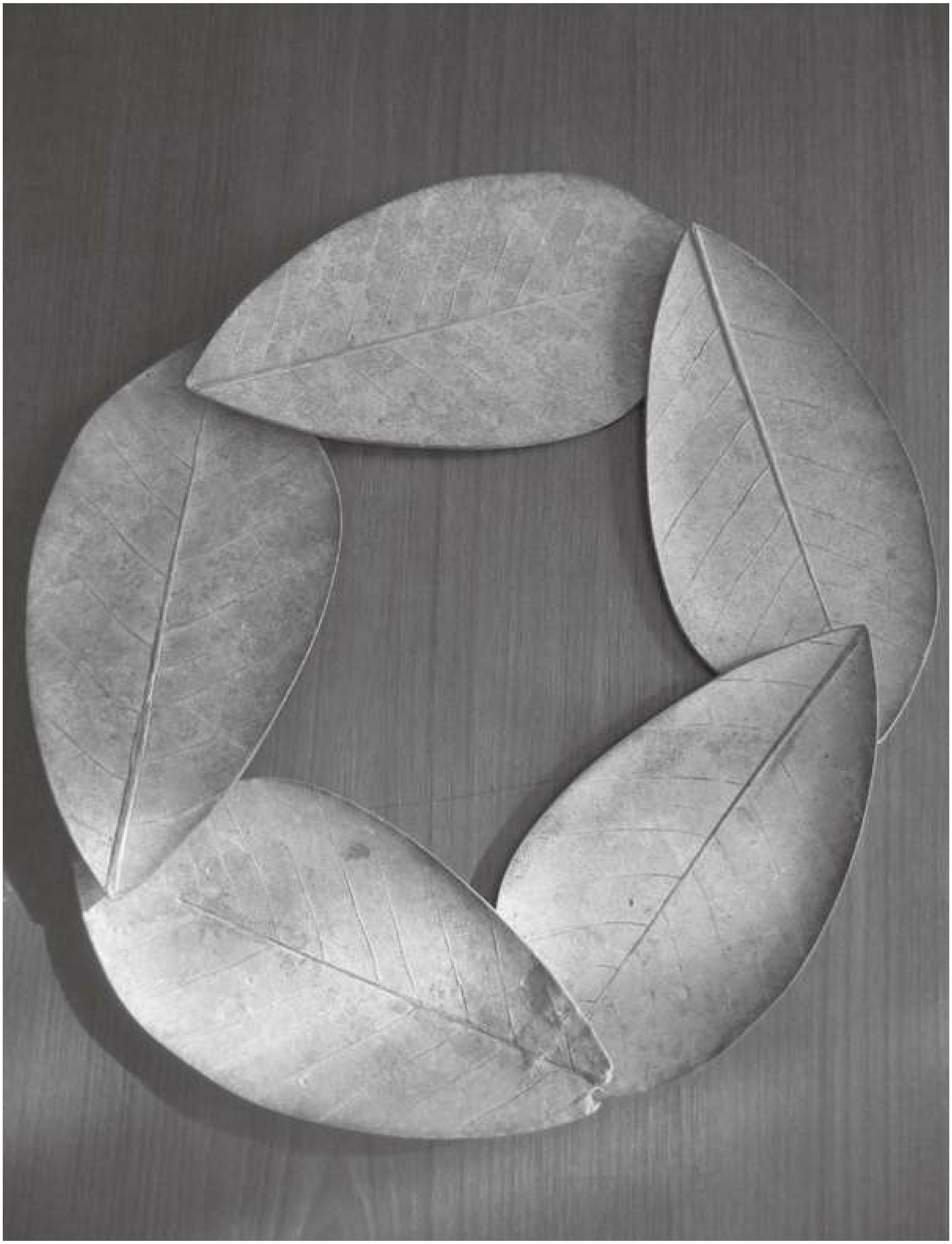
对于我在斯坦福大学的那场视频旅行来说，为预料之外的视觉模式起个生动而有益的绰号，显然不在我最初设想的计划方案之中，但是这个小游戏很快就变得必不可少了。最开始的时候，我以为自己手头的工程涉及的都是直白的术语，如"屏幕之中的屏幕""银色条带""倾斜角度""镜头拉近放大"等等——但我很快就发现自己被迫地、无可奈何地开始使用完全出乎意料的词汇，来描述我观察到的东西。就像你已经看到的那样，我开始谈论起"走廊"和"墙"、"廊门"和"星系"、"螺旋"与"黑洞"、"轮轴"与"轮辐"、"心瓣"与"脉搏"，诸如此类。在我和比尔一起的第二次视频旅行中，这些同样的术语再次成为需要，还有一些新词应求而生，比如"海星"、"奶酪"、"火"、"泡沫"及其他。

在我刚刚产生视频反馈这一观念时，绝想不到自己会跟上面这些词打交道。虽然我是把这些术语应用在一个机械化的确定的系统之上，但

是作为循环的结果而呈展出来的模式却是难以预料的。正因如此，这里需要什么样的词汇，果然也没有人能够提前预知。

如“走廊”、“星系”以及其他种种简单而又诱人深思的隐喻，结果在描述我在屏幕上目睹的那些抽象形状与事件时，成为必不可少的工具。我最初曾在心里默默以为自己会使用的术语，最后基本上都被忽略不提了，因为它们实在给不出什么启发。当然，从原则上说，以它们那种严谨而难于理解的冗长方式，这些术语可以解释一切（就像写出如阿伏伽德罗常数那么多的方程式来解释一种气体的温度和气压一样）——但是这种近乎解析逐个像素的如此无聊的还原主义解释，将会完全错失更高层级的视觉现象，而那种美妙的现象才真正能够引发人类肉眼与思维在直觉上的共鸣。

简而言之，环路模式可以引发意想不到的新结构，进而构成一种新的现实层级，其在原则上可以从基本的环路及其细节化的属性中推导出来，但在实践中却拥有一种不同的“属于它们自己的生命”，呼吁——至少对于我们这些极其有限、追求简单、热爱模式的生物而言——一套新的词汇表和一种新的描述等级，并超越它们从中呈展出来的那个基本等级。



1. 与此相关的一个有趣的例子是关于我儿子的：那是在一个圣诞节，我和孩子飞往加

利福尼亚州。当我们降落到圣何塞机场时，我儿子丹尼盯着窗外问我：“你猜我看到啥了？”“不知道。”我没猜到。他接着说：“停车场里好多车子的头灯和尾灯都一闪一闪的。”我问他它们为什么一闪一闪，他说：“那是因为一辆车子的警报系统一响，就会引发另一辆车子也响。以前在放烟花时，我也见过的。”

2. See [Peitgen and Richter].

3. 浩渺的蓝色远方（wild blue yonder），指神秘遥远的未知地带，也是美国空军军歌的标题。作者把蓝色换成红色或黄色，是为了说明细节参数在这里并不重要。——译者注

第6章
关于自我与符号

“我”的胚芽的感知环路

我饶有兴致地发现，除了专有名词和专有形容词以外，英语这门语言中唯一总是大写的一个词，就是第一人称（主格）代词，也正是我这句话英文原文的华丽开场。这个惯例醒目而怪异，暗示出这个词的所指一定是某种非常重要之物。的确如此，对于某些人——可能是大多数或者甚至是我们所有人——而言，作为“我”或“第一人称”那妙不可言的感觉、关于“在那里”或者仅仅“存在”的直观感受、“拥有经验”与“拥有原始感觉”带来的有力感[注]，似乎是他们生命中最为真实之物，而任何一种主张如果提出，以上所有这一切都可能是一种错觉，或者其仅仅是发生在“第三人称”（即非生命的）事物之间某种物理过程的产物，都会激起一个坚决的内心声音的狂怒。我在此的目标，正是要与这种刺耳的内心声音展开论战。

我先从一个简单的事实说起：对于由进化决定了形态的生物来说，最为根本、天然和内置的目标就是生存。为了提高生存的概率，任何生物都必须有能力灵活应对发生在其所处环境中的事件。这就意味着，不管多么初级和原始，生物必须对其所处环境中发生之事发展出感知和归类的能力（绝大多数地球上的生物都可以安然地无视撞向木星的彗星）。然而，这种感知外部发生之事的能力一旦得到发展，一种奇妙的副作用也会继而发生，并将带来生死攸关的根本性影响。这个事实就是，生物感知环境中某些方面的能力会翻转过来，赋予生物感知关于其

自己的某些方面的能力。

这种翻转的发生，丝毫也不引人惊叹或惊呼不可思议；恰恰相反，它只是与生物的知觉能力伴随而生的一种相当不起眼的副产物，着实微不足道。这就跟音频反馈可以发生一样，稀松平常；或者跟电视摄像头可以指向屏幕并把自身的图像传送上去一样，没什么好惊讶的。有些人可能觉得这种自我感知的概念是古怪的、无所谓的，甚至是无理取闹，但是在这样一种偏见中，自我感知并不是一种复杂或微妙的观念，更不用说是自相矛盾的了。毕竟，在为了生存而挣扎的情况下，永远存在于它所处环境中的一件东西就是……它自己。所以，在所有的事物当中，凭什么单单对在它的世界里最为显著的东西视而不见呢？如此看起，那样才是真正的无理取闹吧！

那种感知上的豁口，让人想起一种语言，它的词汇库不断地增长、扩充，却从不开发出任何单词，用以表达由下面这些汉语词汇所命名的常见概念："说""讲""词语""语言""理解""问""问题""回答""谈论""交谈""声称""否认""争论""讲述""句子""理论""故事""书""阅读""坚持""描述""翻译""转述""重复""谎言""回避""名词""动词""拼音""笔画""声调""偏旁""意义""语法""强调""提及""发音""夸大""恐吓"等等。如果这样一种无视自身的语言真正存在，那么随着它的灵活性和精密性逐渐提高，讲这门语言的人就会展开越来越多的交谈、争论、恐吓及诸般类似的行为，同时却从不提及这些动作；而如问题、答案和谎言这些实体，即便仍然没有命名，也会变得越来越丰沛而显著。就像我们从罗素那怯懦的类型论中看到的跛脚的形式主义一样，这种语言在其核心处也有一个敞开的漏洞——它缺乏以一个词、一段话或一本书（等等）来指涉自身的任何机制。以此类比，对于一个生物而言，进化出丰富的感知与归类能力，却从体制上注定无法把任何这类装置聚焦在自身之上，亦是相当不正常。它的这种选择性无视是病态的，也会为它的生存带来威胁。

环路的多种多样

可以肯定的是，最原初的生灵几乎或者完全就没有自我感知。作为类比，我们可以想象一个电视摄像头被牢牢地捆在一台电视机的顶部，背对着电视屏幕，就像一盏紧紧绑在矿工头盔上的探照灯，总是远远地从矿工的双眼出发向外射去，而从不照入双眼之中。显然，如此设置电视机，就根本谈不上什么自行转动的环路了。不管你如何转动，电视机和摄像头总会同步转动，防止环路的闭合。

接下来，让我们想象一个更加“先进”，因此也更加灵活的设置；这一次，摄像头没有被捆在电视机之上，而是用一条“拴狗绳”拴在电视机上。这样一来，随着绳索的长度和弹性发生变化，摄像头也许有可能扭转到让它的取景器捕捉到至少一部分电视屏幕的程度，得到一条截断式的走廊。达到这种繁复程度的反馈，对应于生物，这相当于我们的宠物甚或很小的孩子形成微弱的自我意识的方式。

显而易见，接下来的一个阶段，就是让“拴狗绳”达到足够的长度和弹性，令视频摄像头得以直接指向屏幕的中心。这将使一条无穷无尽的走廊成为可能，远比一条截断式的走廊丰富。即便如此，闭合自我观察之环路的可能性也无法涵盖系统的丰富性，因为这里仍有许多其他敞开的选项。摄像头能否倾斜呢？如果能的话，能倾斜多少？摄像头可以拉近和伸远吗？它的成像是彩色的还是黑白的？亮度和对比度可以微调吗？成像的分辨率有多高？相对于对环境的观察，花在自我观察之上的时间占据了多大的百分比？有办法让视频摄像头本身出现在屏幕上吗？不一而足。可以做文章的参数还有很多，所以那个潜在的环路也拥有很多开放的繁复性维度。

接收VS感知

尽管以上这些选项提供了很大的丰富性，观看自身的电视系统将总是欠缺一个关键的方面：与单纯的接收或图像接收相对的感知的能力。感知的起点是某种由大量微信号组成的输入（可能是二维的图像，但并不一定如此），但是它随后会更进一步，最终在庞大的休眠符号集合中选择性地触发一个小的子集。符号，即拥有表征品质的离散结构。也就是说，大脑内的一个符号，就像那个假设性的闹壳里的小磁交弹球一样，应该被视作一个可触发的物理结构，并构成了大脑执行一种特定范畴或概念的方式。

关于这带有新义的“符号”一词，我应该提供一个简短的预先声明，因为这个词负载了太多先在的意义联系，而其中有一些是我绝对想要极力回避的。我们经常用“符号”指示书写的标记（字母、数字、写在纸上的音符、汉字等等），但这不是我在此所指的意义。我们有时还会把神话、梦境或语言中的对象说成是代表了某种其他东西的“符号”（比如，一把钥匙、一团火焰、一个指环、一把剑、一只鹰、一根雪茄、一条管道），可这也不是我在此所指的意义。我使用“大脑中的一个符号”这个短语，想要传达的观念是，每当你想到某事，好比埃菲尔铁塔吧，在你的大脑（或者你的闹壳，这取决于你属于什么种类）内部的某种特定的结构便会被激活，而那个大脑的结构，不管其可能是什么，我都会将其称为你的“埃菲尔铁塔符号”。

你还会拥有一个“阿尔伯特·爱因斯坦”符号、一个“南极洲”符号和一个“企鹅”符号。最后这一种，就是当你看到一只或多只企鹅，或者甚至当你并没有看到而只是想到了企鹅时，在你的大脑内部被触发的某种结构。在你的大脑里，有表示动作概念的符号，比如“踢”、“吻”和“杀”，还有表示关系概念的符号，比如“之前”、“之后”和“之间”等等。于是，在本书中，大脑中的符号就是与概念相对应的神经实体，这就好比于基因是与遗传特质相对应的化学实体一样。每个符号在大多数的时间里都处于休眠状态（毕竟，我们之中的绝大多数人都绝少想到棉花硬糖、蛋花汤、圣奥古斯丁、费马大定理、木星“大红斑”或者牙线），但是从另

一个方面说，我们大脑集合中的每一个符号都潜在地具有在任何一个时刻被触发的可能性。

由大量接收到的信号抵达一小把被触发的符号，这条通路是一种漏斗式过程。最初输入的信号受到了操纵或“拿捏”，结果选择性地触发了进一步（即更加“内部”）的信号，依次类推。这种由信号小分队展开的接力传递，在大脑中开辟出一条越来越窄的路线，最终触发一个小小的符号子集(注)，而这些符号的作用自然与初始输入信号息息相关。

在此可举一个有趣的例子。一名走入机场大厅的旅行者，鼻孔里颤动着数以亿万计的微型嗅觉细胞，而根据这名旅行者的饥饿状态和过往经历，这种群颤可能联合触发“甜”和“味”这两个符号，也可能触发“蜜糖”和“发胖”这两个符号，可能触发“肉桂卷”和“附近”这两个符号，也可能触发“飘香”、“广告”、“暗示”、“狡猾”和“噱头”这些符号——还有可能在大脑中触发以上全部11个符号，或是以这种次序，或是以那种次序。这些符号触发中的每一种都构成了一个“感知”行为，与仅仅是接收到自源头传来的大量微观信号截然相对，而后者就仿佛是千百万滴雨珠洒落屋顶一般。

出于清晰阐明的目的，我绘制了一幅过于简单的画面，来呈现感知的过程，而在现实中，却存在着大量的双向流动。信号的传送并非单纯是由外向内，朝向符号进发的；过去的经历也在同时自特定的符号而向外散发出信号。在向内与向外的信号之间发生了某种沟通，结果便是锁定了一条连接原始输入与符号阐释的通路。这种大脑中不同方向流动的混合，让感知成为一种真正复杂的过程。但是，就当前的论述目标而言，我们只需要知道，感知即意味着，在一种快速的双向路径信号传递的作用下，输入信号流的冲撞最终触发了一个小小的符号子集，或者用不那么生物学的话说，就是激活了少许概念。

总之，不管视觉保真度有多高，视频系统中还是缺乏一种成分，即

一个可被选择性触发的符号集合。只有当这样一种集合存在且受到访问时，我们才能说系统真的在感知某物。话虽如此，谁也不能阻止我们去想象，可以把一个普通的视频系统扩充升级，配以高度复杂而精密的附加电路，令其得以支持一连串的信息拿捏过程，并导向一个具有潜在可触发性的符号集合。没错，想一想一个人会如何面对这种工程上的挑战，有助于我们同时想见生灵大脑中的感知过程与其在人工智能（或者就此而论，也可以是外星生物）认知系统中的对应物。然而，显而易见的是，所有这类构造的实现形式，不论其为地球生物、外星生物还是人工智能，其享有的在外来刺激下具有潜在可触发性的符号集合，绝非同等丰富。于是，就像我在本书中一直以来所做的那样，我又想把繁复程度再往上调高一点了。

蚊子符号

让我们先从一只谦卑的蚊子开始(注)（这并不代表我认识任何傲慢的蚊子）。这样一种低级生物对于外部世界具有何种表征？换句话说，在它的脑内装有的那套可被感知过程利用的符号集合是什么样的？甚至于说，一只蚊子是否知道或者相信有存在于“外面”的物体呢？尽管我对此保留怀疑的态度，还是让我们先假设答案为“是”。那么，它会把接收到的对象归于某个种类吗？像“知道”或“相信”这些词，会在任何意义上适用于一只蚊子吗？

让我们说得再具体一点。一只蚊子（当然不会使用词语）会把外部世界划分成诸如“椅子”、“窗帘”、“墙壁”、“天花板”、“人”、“狗”、“毛”、“腿”、“头”或“尾巴”之类的思维范畴吗？换句话说，一只蚊子的脑内是否具有代表这些相对较高水平抽象概念的符号，即离散的可触发结构呢？这看起来几乎是不可能的；毕竟，一只蚊子想要完成蚊子该做的事，完全用不着这些“智力”奢侈品也能做到很完

美。谁在乎我正在叮咬的是一只猫、一条狗、一只老鼠还是一个人呢？谁在乎那是一条胳膊、一只耳朵、一条尾巴还是一条腿呢？——反正我是在吸血就对了。

那么，有哪些种类的范畴是一只蚊子需要的呢？在我所料想的范围内，它的分类系统最为丰富的情况，大概也就只包括某种类似“潜在的食物来源”的东西（简称为“好东西”）和某种类似“潜在的着陆地点”的东西（简称为“港口”）了。它也许还能微弱地意识到某种被我们人类称为“潜在威胁”之物——某种特定的快速移动的暗影或者视差（简称为“坏东西”）。但又要说了，即便加上了“微弱”这个修饰语，“意识”这个词还是用的太重了。这里的关键问题是，一只蚊子究竟是拥有代表这些范畴的符号，还是恰恰相反，只是在用一种类型更加简单的机制蒙混过关，其实并没有能够触发符号的任何感知性的信号级联。

绕过符号，并代之以一种十分精简的感知替代物——如果这种说法让你感到略显模糊，不妨想一想下面这些问题。不管程度多么微弱也好，一个马桶能够意识到它的水位线吗？不论多么稀薄，一个恒温调节器能够拥有对其所控温度的意识吗？就算再细微也好，一颗跟踪导弹能够意识到从它所追逐的飞机身上散发的热量吗？哪怕只停留在最初级的层面，探索科学博物馆里的那个轻盈跳跃的红点能意识到总被它欢快地躲开的人们吗？如果你对这些问题的回答都是“不能”，那么便可以类似地想象一只蚊子的脑袋里拥有一种不需要意识参与的机制，这种机制能让它找到血源、躲避拍打，但是在完成这些壮举时，并没有用到任何观念。

蚊子自己

考虑过蚊子符号之后，我们下面可以向我们探求的核心更近一步了。一只蚊子的内在性的本质是什么？也就是说，一只蚊子对于“我”的

体验是什么样的？一只蚊子具有多么丰富的自我感？这些问题的野心太大了，我们不如试试简单一点的问题。一只蚊子拥有对自己外观的视觉意象吗？但愿你能分享对于这一点所持的怀疑的态度。一只蚊子知道自己长有翅膀、腿或者脑袋吗？它上哪儿才能学会“翅膀”或“脑袋”这样的观念呢？它知道自己长有眼睛或长喙吗？这仅仅是说出来，就已经显得荒诞不经了。它怎么可能搞清楚这种事呢？让我们调整方向，再略微谈谈我们的蚊子对于自己内在状态的知识。它知冷暖吗？有精力充沛或疲惫感吗？能意识到自己的饥饿难耐吗？能分清快乐和悲伤吗？希望或恐惧呢？我很抱歉，但即使是这些感觉，就我的感受而言，对于如蚊子这般卑微的实体，都是不切实际的。

好吧，那么像“疼”或“不疼”这样更基本的感觉呢？我仍然持怀疑的态度。而从另一个方面来说，我很容易便可想象出，蚊子眼睛里传出的信号发送到它的大脑，引起其他信号反弹回它的翅膀之上，形成一种反射，而这种反射可以被我们人类以言语表达为“左边危险快逃”或者仅仅是“滚开！”——但是我担心的是，即便以这种电报方式的措辞，仍然会让蚊子听起来太有自觉性了。我很乐意把一只蚊子的内在生命与一个冲水马桶或者一个恒温控制器的内在相提并论，但这大约就是我个人所能做到的极限了。蚊子的行为带给我的印象是，无须诉诸任何称得上“符号”之名的东西，也完全可以理解。换句话说，一只蚊子没有言语也没有概念的规避危险的行为可能不太像我们人类所熟悉的感知，反而更像是没有言语也没有概念的膝跳反射——当你的医生拿锤子敲打你的膝盖时，你就会发射性地抬腿一踢。一只蚊子拥有比你的膝盖更多的内在生命吗？

一只蚊子对于自己有没有哪怕是最为微弱的隐约一瞥，窥见自己是在一个庞大的世界之内的一个移动的部分？一如既往，我怀疑是没有的，因为这就要求在它微型的大脑里容下各式各样的抽象符号——表示“大”“小”“部分”“地点”“移动”等等观念的符号，更不用提“我自己”了。一只蚊子为何需要如此奢侈的东西呢？它们如何能帮助蚊子更

加有效地找到血源或者配偶呢？假设有一只蚊子，它拥有足够强大的脑力，可以容纳这些奇妙的符号，那么它就会成为一个头重脚轻的蚊子学究，脑子里随身装着比自己那些头脑更简单、结构更简化的兄弟姐妹们远远多出的神经元，因而也会比它们更加笨重和缓慢，也就意味着它没有能力与它们在寻求血源与繁殖的方面竞争，从而在进化的赛跑中遭到淘汰。

无论如何，我的直觉是，在一只蚊子十分有效率的微型的神经系统中，完全不存在感知分类的范畴（以及符号）。如果我说的没错，那么这就把可以存在于蚊子大脑中的那种自我感知环路降到了一个极低的水平，也让蚊子在事实上成为一种特别“小灵魂的人”。如果我提出，一只蚊子的“灵魂”大概与在探索科学博物馆那面墙上跳来跳去的红色光点拥有差不多的“尺寸”——比方说，100亿分之一亨内克（即大约是人类灵魂的1万亿分之一），我希望这听起来不会太过亵渎或疯狂。

不可否认的是，我在估计这个数值时，是很轻率的；但是，我在展现自己关于蚊子大脑中是否存在符号的这一主观猜测时，却是相当严肃的。尽管如此，这仍然是一种主观的猜测，而且你也大可不必赞同，但是有关这种细枝末节的异议，在此并非紧要。真正关键的要点，远比这更加简单和粗略，仅仅是：存在某一种生物，在根本上只适用于这种等级的复杂性，而不能更多。如果你不赞成我的这个论断，那么我想邀请你在那个由多种多样的动物智力所构成的标尺上，来回上下滑动一番，直到你觉得自己卡中了那个适当的等级为止。

对所有这些再做最后一点反思。有些读者可能会带着一种听起来诚挚万分的腔调，抗议所有那些有关一只蚊子眼中的世界的问题：“我们怎么可能知道呢？不论是你还是我，都不可能进入一只蚊子的大脑或思维——没人可以。据我所知，蚊子是和我一样有意识的！”好吧，恕我直言，这种声明不可能是真诚的，因为我愿意跟你赌10美元，押这些读者都会不假思索地拍死一只停在他们胳膊上的蚊子。如果他们真心相信

蚊子有可能与他们自己拥有一样的感知力，那么他们怎么可能愿意于须臾之间扼杀蚊子的生命呢？如果真如他们所说，这些生命很可能与人类享有同样多的意识，那么他们还能在处死这些生灵时不多眨一眼，难不成他们都是穷凶极恶的怪物吗？我觉得要想判断一个人的观点，你不得不依据他们的行径，而不是他们所说的话。

一则关于机器人汽车的插叙

在向更高级的动物种类进发前，我想插入一段关于机器人汽车的简短讨论，这些汽车可以在平滑的公路上或者多石的沙漠中自行驾驶前进注。这种车辆上都安装着一台或多台电视摄像机（以及激光测距仪和其他种类的传感器），配合特别的处理器，令车辆得以掌握周围的环境。仅仅对屏幕上的颜色或原始形状进行的简单分析，无论多少，都无法给出绕过障碍而不至于翻车或者被卡住的有益建议。为了成功地自行驾驶，这样一种系统必须拥有一套规模不小的仓储库，存入预先打包好的可被外部场景选择性触发的知识结构。于是，机器人汽车要想避免陷入泥潭、困于隘谷或是卡在两块巨石之间，就需要关于某些抽象概念的知识，如“道路”“山丘”“隘谷”“泥泽”“岩石”“树木”“沙子”以及许多其他的概念。电视摄像机和测距仪（等）只服务于车辆“感知过程”中最为简单的初始阶段，而触发上述诸般不同的知识结构，才对应着过程的远端，即符号的一段。

关于是否要在上一句话里的“感知过程”几个字上加双引号，我曾略微迟疑过，但还是做出一个任性的选择，心想不管加还是不加，我反正都讨不到好处了。意思是，如果不加引号，我就是在暗示，在这样一辆机器人汽车的视觉输入处理过程中，发生了与我们自己的感知切实一致的事情；而如果我加了引号，就是在暗示，在“单纯的机器”可以做的事情和生灵所做的事情之间，横亘着某种不可逾越的鸿沟。每种选择都代

表了一种太过非黑即白的立场。遗憾的是，引号本来就不提供灰度；如果它们提供的话，我倒想使用中间的某一灰度，来标示出一种差别更加细微的立场。

如今的自我导航机器人汽车尽管已经足够令人叹服，但是距离哺乳动物的感知水平仍然相当遥远，可我依然觉得可以不失公允地说，这种汽车对于其所处环境的“感知”，（再次为没有灰度的引号道歉！）在繁复程度上正与一只蚊子的“感知”（这里也用引号了——我希望能多少找回一点平衡）相当，而且可能更加复杂。瓦伦蒂诺·布赖滕贝格（Valentino Braitenberg）在他所著的《车》（*Vehicles*）这本书中，对机器人汽车这个概念以及不同等级的“感知”给这些汽车带来的收益进行了美妙的处理。

不必追究更多的细节了，请允许我有话直说。在本书中，在同一个部分里讨论有生命的动物和能够自我引导的机器人，是完全合理的；因为，当今的科技成就正让我们前所未有地接近理解在那些存活于复杂环境里的生命系统之中所发生之事。这些成果拆穿了约翰·塞尔口中喋喋不休的陈腐邪说，说计算机永远只能是生命过程的“模拟”。如果一辆自动车能够驾驶自己达200英里之远横穿一片令人望而生畏的沙漠地带，那么这项壮举怎么可能仅仅被称为一种“模拟”呢？这当然也是应对恶劣环境的一种求生行为，同一只蚊子在房间里飞旋并躲闪拍打的行为一般真切。

细思狗想

让我们回到纯生物的感知繁复度阶梯之上，沿着从病毒到细菌、蚊子、青蛙、狗、再到人的次序继续向上攀登（我知道，我在中间跳过了几级）。随着我们转向越来越高的水平，可触发的符号集合自然也会越来越丰富——真的，不然“沿着阶梯向上爬”还有什么另外的意思呢？仅

仅从宠物狗的行为上判断，没有人会怀疑，它们已经发展出一套可观的范畴集合，包括且不限于以下的例子：“我的爪子”“我的尾巴”“我的食物”“我的水”“我的碟子”“屋里”“室外”“狗门”“人走的门”“敞开”“关闭”“热”“冷”“夜晚”“白天”“人行道”“马路”“灌木丛”“草地”“拴狗绳”“散步”“公园”“汽车”“车门”“我的大主人”“我的小主人”“猫”“友好的邻家的狗”“凶恶的邻家的狗”“快递货车”“兽医”“球”“吃”“舔”“喝”“玩”“坐”“沙发”“爬上”“不良举止”“惩罚”等等。导盲犬通常要学会100个或更多的词汇，并能在很多不同的语境中对这些概念千变万化的表达形态做出反应，因而展现出它们内在范畴系统（即它们的可触发的符号集合）的丰富性一面。

为了表现一种犬类范畴集合的本性，我使用了一套人类语言的词汇，这当然不是在说，当一条狗对邻家的狗或者快递货车做出反应时会牵涉人类的语词。但是有一个词特别值得一提，就是在“我的尾巴”或“我的碟子”中出现的“我的”。我怀疑大多数读者都会同意，一只宠物狗能认出某只特定的爪子是属于它自己的，而并非仅仅是环境中一个随机的事物或者某个其他动物身上的部位。同样，当一条狗追逐自己的尾巴时，即便它不可能意识到这一行为的反讽意味，但它也一定知道，那条尾巴是它自己身体的一部分。由此，我实际上是在提出，一条狗具有某种基本的自我模型，也具有某种自我意识。除了有表示“汽车”、“球”和“拴狗绳”的符号和表示其他动物和人类的符号之外，它还拥有某种表征其自身（指这只狗自身，而非符号自身！）的内在大脑结构。[注]

如果你怀疑狗不具备这些，那么黑猩猩怎么样？两岁的人类婴儿呢？不管怎样，这样一种反射性符号结构的出现，无论其最开始显现的感知力的层级如何，都构成了“我”的核心胚芽和最初的火花，而在这微小的核心之上，更加复杂的对“我”的认识，会在接下来的一生中不断生长，就像雪花的生长是从微小的初始尘粒开始一样。

鉴于几乎所有成年的狗都有一个表示狗的符号，那么一条狗会在某种意义上知道它自己也属于狗这一范畴吗？当它注视一面镜子，看到自己的主人站在“某只狗”的旁边时，它能认出那只狗就是自己吗？这些都是十分有趣的问题，但是我不会在此尝试给出答案。我怀疑，这种认识能力游走在犬类思维能力的边缘，但是至于狗到底落在边缘的哪一边，对于我这段文字的目的而言，其实是无关紧要的。说到底，这本书原本也不是关于狗的。这里的关键要点在于，存在某种复杂性的等级，令生物可以开始把某些范畴应用在自己身上，开始搭建表征自身的思维结构，并开始以某种“知识视角”把自己置于与世界其余部分的关系当中。就这方面而言，我认为狗要比蚊子高级很多，而且我猜你并不会表示反对。

而另一方面，如果我说一条狗的灵魂显著地“小于”一个人类的灵魂，我猜你也不会反对——不然的话，为什么你和我都没有跑到各自所在地的宠物收容所去强烈地抗议示威，抵制那里每天把流浪的猎犬和走失的小狗处以“人道毁灭”的行为呢？你能宽恕把无家可归者和弃婴处死的行径吗？是什么让你在狗与人之间画出了那条区分线呢？会不会是二者灵魂尺寸的相对大小呢？要想让你下定决心，前往动物收容所去组织一场抗议示威活动，平均而言，狗必须有多少亨内克的灵魂才够呢？

达到狗这一繁复水平的生物，得益于其感知器官不可避免的来回翻转及其低调却不容忽视的范畴集合，不由自主地发展出一种对于它们自己的大致的认识，将自己视为在一个更大的世界之中的物理实体。（参加沙漠越野赛的机器人汽车不会把它们宝贵的时间花在打量自己之上——这就跟空转车轮一样没用，所以它们的自我意识远不如一条狗那么繁复。）虽然一条狗永远都不会拥有有关它的肾脏或者大脑皮层的任何知识，但它能发展出某些有关它的爪子、嘴和尾巴的概念，也许还有它的牙齿或舌头。它也许在一面镜子里看到过自己，并认出“那只贴在我主人旁边的狗”实际上是它自己。或者它也许和主人一起在家庭录像中看到了自己，识别出了它主人被录制下来的声音，而且认出了视频里的

叫声来自它自己。

然而所有以上这一切，虽然在很多方面都相当了不起，但是与一个正常人类毕生持续发展的那种对自我和“我”的认识相比，仍旧是极其局限的。为什么会这样？在二黄、路飞、黑子、小斑和老八公的身上，到底缺失了什么？(注)

截然不同的人类概念集合

进化在人类这一点上裂开了一道惊人的鸿沟，而又渐渐地与其他灵长类动物越隔越远：它们的范畴系统变得可以任意扩展。一种抓人眼球的开放性品质走入了我们的思维生活，其在根本上是一种没有限制的可扩展性，与其他物种十分常见的有限性形成了鲜明的对比。

人脑中的概念具备了一种属性，令它们得以与其他概念糅合在一起，形成更大的概念包，而这种更大的概念包本身，随后又会成为一个新的概念。换言之，概念可以等级化地相互嵌套于彼此之中，而这种嵌套可以发展到任意的程度之上(注)。这不免让我想起——且我不认为这纯属偶然——在视频反馈里，一条无穷无尽的长廊与一段截断式的走廊之间所存在的巨大差异。

举例来说，繁衍后代的现象可以引发出一些概念，如“母亲”、“父亲”和“孩子”。这些概念又引发出“父母”这一嵌套式的概念——说它是嵌套的，是因为这一概念的形成依赖于三个先在的概念：“母亲”、“父亲”和“非此即彼”这一抽象观念。（狗会具有“非此即彼”的概念吗？蚊子呢？）一旦“父母”这个概念诞生，又会为“（外）祖母”（父母的母亲）和“孙辈”（孩子的孩子）的概念打开大门，随后便是“曾（外）祖母”和“曾孙辈”。我们形成的所有这些概念都是蒙嵌套所赐。当加入了“姐妹”和“兄弟”的概念之后，拥有更高嵌套等级的概念，如“叔

舅”、“姑姨”和“（堂）表兄弟姐妹”等，将随之出现。接下来，还会有嵌套程度更高的概念发生，比如“家庭”。（“家庭”的嵌套程度更高，是因为它把所有以上这些先在概念视为理所当然，并建立于它们的基础之上。）

在集体性的人类观念圈内，概念的建构通过这种合成式的行为开始滚雪球，而不知局限为何物。人类这个物种很快就发现自己竞相向上，跳跃至更高的概念之上，如“风流韵事”“三角恋爱”“忠诚”“诱惑”“复仇”“绝望”“疯癫”“精神崩溃”“幻觉”“错觉”“现实”“幻想”“抽象”“梦”以及当仁不让地占据世间最高峰顶的“电视剧”（这里面还会嵌入“广告插播”、“牙好，胃口就好”以及“X品牌”这些概念）。

让我们想想“食品杂货店收银台”这个看起来稀松平常的概念。我敢打赌，这一定是你个人概念集合里的一名合格成员。它听起来就像是一个嵌套式的实体，由好几个词组合而成，而这些组成的词由此直白地告诉我们，它所代表的是买卖食品杂货的商店里结账的窗口。但是仅仅看到外显的词汇结构，无异于蜻蜓点水。实际上，这个概念中还包括了很多很多的其他概念，列举如下：“购物车”“排队”“顾客”“等待”“糖果架”“硬糖”“小报”“电影明星”“八卦头条”“肮脏的丑闻”“每周电视节目表”“肥皂剧”“青少年”“围裙”“座签”“收银员”“心不在焉地打招呼”“收银机”“键盘”“价格”“数字”“附加”“扫描机”“二维码”“哔哔”“激光”“滚动带”“冷冻食品”“锡罐”“蔬菜包”“重量”“标尺”“折扣券”“橡胶分隔栏”“滑动”“装袋工”“纸袋”“塑料袋”“塑料卡”“现金”“装货”“付款”“信用卡”“借记卡”“刷卡”“收据”“圆珠笔”“签名”等等。这张清单开始看上去无穷无尽了，而我们正在讨论的内在丰富性，还只不过属于一种极其平常的人类概念。

当然，我们在想起一个食品杂货店收银台时，并不需要激活所有这些概念成分——这里不仅存在一套无论如何总能被激活的概念核心，还有很多可能不被激活的更加边缘的成分，但是所有以上这些，以及显然

还有更多的概念，共同构成了我们思维中的一个完整概念。另外，这个概念与我们思维中另外的所有概念都一样，能够完美地被收纳入其他的概念之中，比如“食品杂货店收银台浪漫史”或“玩具食品杂货店收银台”。你大可围绕这个主题尽情发挥，创造出属于你自己的变体。

情景记忆

当我们和朋友一起围坐在桌子上侃大山时，难免会想起过去某个时候发生在我们身上的情景，而且通常是很多年以前的事。那一次，我的狗在街区里走丢了；那一次，我们邻居家的孩子在机场走丢了；那一次，我们只晚了一步，没赶上飞机；那一次，我们赶上了火车，但是我们的朋友晚到了一步；那一次，火车里闷热得要命，我们不得不一路站在过道里，足足四个小时；那一次，我们上错了火车，直到一个半小时之后才下来；那一次，所有人的口中都只能说出一个单词：“玛——丽——莲·梦——露”，说的时候还以一种诡异的嬉笑姿态凭空比画出一个沙漏的形状来；那一次，我们开着车，在斯洛文尼亚的午夜彻底迷失了方向，几乎用光了汽油，可竟然凭借所掌握的屈指可数的斯洛文尼亚单词找到了开往意大利边境的路。如此可以一直回忆下去。

情境勉强称得上是一种概念，但是它们是在时间的推移中发生的，都是独一无二的，有点像是专有名词，却没有一个名字，而且与时间中的某个特定时刻联系在一起。虽然每一个情景都是“唯一”的，但是它们也会落入各自所属的范畴，上面那一段文字已经使出“你懂的”眼色暗示出了这一点。（差一步没赶上飞机并不是独一无二的事情，即使这在你的人生中只发生过一次，你也极有可能听说过好几个属于这一范畴的成员，而且能够轻易地想象出数不尽的其他范例。）

情景记忆是我们私人的情景仓库，而这些情景都曾发生在我们自己、我们的朋友、我们看过的小说、电影、新闻故事或者电视新闻片段

中人物的身上，形成了长时记忆的一个主要部分，并令我们如此具有人情味。显而易见，关于情景的记忆可以被我们目睹的外部事件或者其他已被触发的情景所触发[注]，而且同样显而易见的是，几乎所有关于特别情景的记忆在绝大多数的时间里都处于休眠状态（不然的话，我们会彻底疯掉）。

猫或狗也有情景记忆吗？它们能记住几年、几个月、昨天乃至于几分钟之前发生的特别的事件吗？当我带着我的奥利出门跑步的时候，它能回忆起自己昨天是如何一边被拴狗绳往回拖着、一边努力地朝街对面那只（也被拴狗绳绑住的）达尔马提亚狗说“嗨”的吗？它能记住我们三天前走过一条与往日不同的路线吗？当我把奥利送到狗舍去度过感恩节假期时，它似乎记得狗舍是一个地方，但是它记得上一次（或者任何一次）在那里时发生的一件特别的事吗？如果一条狗害怕一个特殊的地方，是因为它回忆起在那里发生过的一次特别的创伤事件，还是因为那个地方就是和某种不祥的感觉联系在了一起呢？

虽然这些问题在我看来都相当奇妙有趣，但我在此并不需要答案。我不是在写一篇论述动物意识的学术论文。我只是想让读者思考一下这些问题，然后同意我的说法，即它们之中有些问题应该得到的答案为“是”，有些问题应该得到的答案为“否”，而有些问题就是不能简单地给出答案。但是，我的大体观点是，我们人类与其他的动物不一样，拥有所有这些种类的记忆；没错，我们拥有的是全套。我们可以历历在目地回忆起15年或20年前假期中的某个特定的情景。我们准确地知道自己为什么害怕某个地方或者某个人。我们可以细节化地还原自己在威尼斯、巴黎或伦敦与某人出乎意料的偶遇。人类记忆的深度和复杂度，其丰富性是令人难以置信的。因此也难怪，当拥有如此丰富的概念和记忆的人类，依其不可避免的必然性，把注意力转向自身时，会产出一种具有超凡的深度与缠结性的自我模型。而这一深刻而缠结的自我模型，正是与“我”相关的全部。[注]

1. 某些哲学家称其为“感质”（Qualia）。

2. See [Kanerva] and [Hofstadter and FARG].

3. See [Griffin] and [Wynne]. The latter contains a remarkable account of analogy-making by bees, of all creatures!

4. See [Davis 2006].

5. 这听起来像是一句玩笑，但其实不完全是。当说到人类的自我符号——他们的“我”——时，这种“我”的大部分结构都包括指向抽象的“我”的指针，而不仅仅是指向身体。这在第13章和第16章中有所讨论。

6. 二黄、路飞、黑子、小斑和老八公，都是狗的常见名字，而且都曾出现在美国的电视节目当中，如“老八公”便出自著名的电影《忠犬八公的故事》。作者列举这些有名字的狗，是为了强调狗与人之间灵魂尺寸具有相对的大小。——译者注

7. I defend this point of view in [Hofstadter 2001]. For more on human categories, see [Sander], [Margolis], [Minsky 1986], [Schank], [Aitchison], [Fauconnier], [Hofstadter 1997], and [Centner *et al.*].

8. See [Kanerva], [Schank], and [Sander].

9. See [Dennett 1991], [Metzinger], [Homey 1942], [Homey 1945], [Wheelis], [Norre minders], and [Kent].

第7章
副现象

要多真实，有多真实

感知的下漏过程终将在以0.001秒为单位的时间里引发动物脑中特定离散符号的激活，得益于此，动物（让我们忘掉机器人汽车吧！）才能够与其所处的物理环境形成亲密与可靠的关系。一个成年人不仅可以轻而易举地剥掉香蕉皮并且避免让自己撞入荆棘丛生的玫瑰丛里，而且还能在瞬息之间，对强烈的气味、奇怪的口音、可爱的娃娃、发出巨响的撞车、吸引眼球的头条、技术非凡的滑雪者、雍容华贵的衣裳等等做出反应。它甚至偶尔还能击中以每小时80英里的速度冲过来的曲线球。因为动物在自身内部对于世界的镜照（mirror）是高度可靠的（大象这个符号不应该被蚊子的嗡嗡声所触发，而蚊子的符号也不应该被眼前踱步而来的大象所触发），因而动物经由自己私有的符号而反映出的世界镜像，便成为一种不可置疑的稳定性支柱。它所感知到的事物与模式定义了它的现实——但是并非所有感知到的事物和模式在它看来都同等真实。

当然，在不会说话的动物之间，不论是明说还是暗指，都从来没有提出过“在我感知到的所有事物里，哪些是在我看来是最真实的？”这个问题。但是在人类的生命里，关于何为真实而何为不真实的这些问题，早晚会不可避免地冒出来，有时候被有意识地小心翼翼地说出口，有时候则缄默于心，尚未成熟，只是在暗地里静静地酝酿发酵。在儿童和青少年时期，我们或亲眼所见，或在电视上看到、在书本上读到，或听他

人讲述了很多理应存在之物，这些事物于彼此之间展开激烈的竞争，争夺着我们的注意，并期望获得我们心中现实性评价标准的认可——例如，上帝、哥斯拉、歌帝梵、戈多、哥德尔、神、女神、鬼魂、食尸鬼、妖精、小妖怪、丑怪、魔像、狮鹫、半鹫半狮、胶子和圣诞怪杰。一个孩子要想理清这其中某些概念的真实性，得花上几年的工夫；实际上，很多人要用一生的时间才能理清（偶尔还需要更长一点的时间）。

我所谓“理清X的真实性”，指的是得出一个稳定的结论，确认你对X的信任度有多大，以及你在对自己和他人解释事物时，能否放心地以X的概念为依托。如果你愿意借用狮鹫来展开你的解释，而且当他人如此做的时候，你也不会敬而远之，那么这样看起来，狮鹫对于你而言便是一个正经而真实的概念了。如果你已经基本上理清了狮鹫对于你自己的现实性，又听说有一期关于狮鹫的特别电视节目将要播出，那么你就不会觉得有必要收看这期节目并以此帮助自己判断狮鹫是否真的存在了。你可能很坚定地相信狮鹫的存在，也可能认为它们不过是一种孩子的幻想或者一个玩笑——不管是哪一种，你的心意已定。又或者，你可能还没有理清狮鹫的现实性；这样的话，如果它在一场晚宴聊天中被人谈及，你则会感觉到无措、困惑、惘然、怀疑或者举棋不定。

另一种思考“X对于你而言有多么真实”的方式，是看你对于一篇把X（例如，一只活着的恐龙、碰见希特勒、火星上发现的昆虫、一台永动机、UFO绑架事件、上帝的全知全能、灵魂出窍的体验、平行宇宙、超弦、夸克、大脚怪、“老大哥”、宇宙大爆炸、亚特兰蒂斯、诺克斯堡的黄金、南极点、冷聚变、爱因斯坦的舌头、霍尔顿·考尔菲德的头脑、比尔·盖茨的支票簿或者众所周知的那堵马拉松长跑者必须跨越的20英里处的“墙”）的存在视为当然的报刊文章有多少信任。如果你一看到X在文章中被理所当然地默认为存在，便中断了阅读，那么看起来X的“真实性”在你这里就是高度可疑的了。

随便挑选一个上面提到的概念。几乎可以肯定的是，有很多人热忱

地相信它的存在，另外一些人则半信半疑，还有一些人完全不信（不论是出于无知、犬儒、不良的教育还是卓越的学识）。这些概念中的某一些，权威专家们已经反复告诉过我们，它们是不真实的，可是我们依然在电视节目、书籍和报刊中一遍又一遍地听到它们，所以我们对于它们是否存在或者是否可能存在的问题，便只剩下一团模糊的好奇感。另外一些概念，权威专家们告诉我们，它们是绝对真实的，但是我们就是从来没有见过它们。还有一些概念，我们听说它们曾经是真实的，而现在已经不再真实了，这就把它们安插在了真实性的某种中间状态之上。而还有一些概念，我们听说它们是真实的，但是完全超越了我们想象的能力。另外还有一些据称是真实的概念，只是在隐喻的层面上成立，或者大约如此——不一而足。要想把这些情况全部理清，可绝非易事。

具体的墙与抽象天花板

把所有这一切归结到一个更加具体的问题上来：前面提到的马拉松长跑者的20英里之处的“墙”有多么真实呢？如果你是个马拉松长跑者，你几乎肯定已经充分地考虑过这个问题了。你可能亲身体验过它，或者认识亲身体验过的人。或者，你也可能认为这是一个夸大其词的概念。我自己从来没有“撞过这堵墙”，但是话说回来，我跑过的最长距离也不过15英里。我知道的是，“他们说”，大多数的跑步者，如果没有接受恰当的训练，都会在20英里左右的距离“撞上一堵无情的墙”：他们的身体耗尽了全部的糖原，转而开始燃烧脂肪（我听过有人把它描述成“你的身体在吞噬它自己的肌肉”）。它在到来之前毫无预兆，撞见之时则疼痛难当——“就像一头大象从树上掉落在我的肩膀。”马拉松长跑者迪克·比尔兹利（Dick Beardsley）这样说。很多跑步者在触及该点之后，就是无法再继续往前了，只好中途退出。但是，这是一种普遍的现象吗？它对所有人都一视同仁吗？会不会有某些马拉松长跑者从来就没有过这种体验呢？而且，即便它在科学上是可以解释的，我们能说它是一种跟

一个人一头撞上一堵实际的墙一样真实而可触的现象吗？

1966年，我初入伯克利，攻读数学专业的研究生。那时，我对自己的印象还是一个相当了得的数学天才。毕竟，我在斯坦福读数学专业的本科时，不但面对几乎所有的课程都不费吹灰之力就能顺利通过，而且还开展了大量的原创性研究，并在毕业时获得了数学系授予的“杰出”荣誉。我希望成为一名数学家，并完成一些伟大的事情。好吧，在伯克利，有两门课是所有一年级学生的必修课——抽象代数和拓扑学，于是我就选上了这两门。令我震惊的是，这两门课都让我感到十分困难——我此前从未遇到过类似的事。虽然我在这两门课上都拿到了高分，但那只是凭借死记硬背之后在期末考试上的照猫画虎。在那整整一年里，我都在为自己意象能力的严重缺乏而头疼万分，这也是我头一遭有如此经历。那就像在攀爬一座高峰的途中，随着空气越来越稀薄而感受到的入骨三分的头痛。抽象叠于抽象之上[注]，而我耕作得越深，我的步伐就越慢，我能把握住的东西也越来越少。最终，在一年半以后，我意识到局面已无可救药；我噙着苦涩的泪水，自信心开始崩溃，终于抛弃了自己的数学家梦想，永远地逃离了这个领域。我在毫无预先警告的情况下，一头撞上了这块可恨而又坚固的“抽象天花板”，这次创伤痛彻心扉，并改变了我的人生。那么如此说来……这块抽象的“抽象天花板”到底是一种多么具体、多么可信、多么真实的东西呢？它和马拉松运动员的“墙”一样真实吗？它和我可以用自己的脑袋撞出声音的木头房梁一样真实吗？究竟何为真正的真实呢？

虽然没有人会刻意如此规划，但我们绝大多数人在度过青春期之后，关于何为真实，都会形成一种充满灰度的洞幽烛微的认识。（然而我也知道，可能读者你也知道，所有那些在我看来微妙莫测的议题，在少数成年人的眼中完全是非黑即白的——完全不必理会任何麻烦凌乱的灰度。这一定能让生活简单不少！）事实上，说对于我们大多数人而言，生活充满了明暗不同的“灰度”，也是太过简单了：这个术语唤起的是一种直白的单一维度连续体的意象，不同等级的灰度游走于白与黑之

间；而事实上，这里涉及的维度不只这么简单。

所有这些讨论都会令人深感不安，因为“真实”这个词和许多词一样，看起来蕴含着一种截然分明、一刀两断的二分法。当然，本来的情况就应该是，有些东西单纯就是真实的，而另外一些单纯就不是真实的。当然，不应该存在任何部分真实的东西——那根本就说不通！可是，虽然我们十分努力地在迫使世界去附和这一理想中非黑即白的二分法，万事万物还是会不遂人愿地模糊成一片。

关于真实性的多面向的知识基础

摆在我书桌上那只小纸板盒里的玻璃球显然是真实的，因为我看到那个纸板盒就放在那里，而且我还可以走过去打开盒子，用手去挤压那颗玻璃球，掂量它，并感觉到它的硬度。我希望你能明白我想表达的意思。

高速公路出口边上那个75英尺高的壳牌加油站指示牌的顶部是真实的，我敢确信，因为每一个指示牌都是固态的，而一个固态的东西总是有一个顶部；同时还因为，我能够看到指示牌的底端和它的两边，所以，通过类推，我也想象自己看到它的顶部；还因为，即便我绝对摸不到它，但我至少可以在理论上爬到它上面或者从一架直升飞机上降落在它的顶部。还有，这个指示牌可能在一场地震中翻到，这样我就能冲过去摸到它的顶部。这种可能性还多的是。

南极洲也是真实的。虽然我从来没去过那儿，而且几乎肯定永远也不会去，但是我看到过不下100张它的照片，还从太空俯拍的地球的照片中知道南极洲的存在，而且我有一次还遇到过一个跟我说自己去过南极洲的人，以及很多其他的理由。

为什么我相信某些人告诉我的事情，要多过相信另外一些人说给我

的话呢？为什么我会相信（某些）照片是真实性的证据呢？为什么我会信任某些特定的图书中的照片？为什么我会信任某家特定的报纸，并只相信到一定的程度呢？为什么我对每一张报纸的信任程度都不一样呢？为什么我不会同等地信任所有的图书出版方呢？为什么我不愿给所有的作者以同等的信任呢？

通过很多不同种类的抽象、类比和归纳推理，通过很多对各种各样权威声音的绵长而曲折的引用链（这构成了一根支撑着每一个成年人信念系统的不可或缺的支柱；不必理会高中老师的顽固论调，他们年复一年地讲授着“权威论断”的虚假性，同时又确信他们自己是应该获得信任的，因为他们说到底正是权威人物），我们围绕着何物存在于“外面那里”，搭建起了一套错综复杂且环环相扣的信仰——然后再一次，不可避免也不容喘息，这套信念又折返回来，用在了我们自己的自我之上。

正如我们相信他人拥有肾脏和大脑一样（这几乎完全是诉诸权威和类比的论断），我们也会相信自己拥有肾脏和大脑。正如我们相信他人具有道德一样（这基本上还是诉诸权威和类比的论断），我们到头来会相信自己也有道德，并且还会相信那张将出现在本地报纸上的自己的讣告的真实性——即便我们明知自己绝不可能亲手翻动那些报纸并阅读自己的讣闻。

是什么令我们对这些如此抽象的东西形成了绝对确信的感觉呢？首先是我们的内在符号直接镜照具体环境的可靠性（例如，我们购买一杯咖啡，则在我们闹壳内部的某个地方瞬时跳出一个反映这杯咖啡的物理记录，从桌面到自己的手里，时刻追踪着它的方位，不断更新它的颜色、苦度、热度和所剩多少）。其次，便是我们的思考机制告知我们的无法用肉眼观察到的那些更加抽象的实体的可靠性（例如，拿破仑在法国历史上的影响、瓦格纳对法国晚期浪漫主义作曲家的影响、由埃瓦里斯特·伽罗瓦等人提出的五次方程的根式不可解[注]）。所有这些更加抽象的东西，都深深地根植于时刻被强化的符号之上，这些符号由我们直

接观察到的世界中发生的事件任意地从休眠状态中激活。而正是这些直接的思维事件，构成了支撑我们更加广泛的真实感的基石。

由此自然可知，在我们看来最为真实的东西，便是被激活得最为频繁的东西。我们手指上的倒刺在我们看来格外地真实（纯属偶然的是，我在重新修改这一段时，刚好察觉到自己在无意地拨弄自己的一根倒刺），但是对于我们大多数人而言，英国的村庄奈勒瓦洛普（Nether Wallop）和喜马拉雅山脉高地上的国家不丹，则明显不如倒刺真实，更不用说在仙女座中缓慢打转的螺旋星系了——即便我们知识上的自我可能更愿意坚持相信，既然后面这些概念比我们手指上的倒刺要强烈许多，也远远更加持久，那么它们对于我们而言，就应该远比倒刺更加真实。我们可以把这些说给自己听，直到口干舌燥为止，但是我们极少有人能在行动中表现出我们是真正相信它的。在某片遥远的大陆上，地下岩石的一次轻微滑脱夺走了20 000人的生命；在亚马孙的盆地里，原始森林无边无际地蔓延，遮天盖地；一群无助的星球，一个接一个地被贪婪的黑洞所吞噬；乃至两个各包含上千亿颗星球的巨大星系，正在上演一场恢宏的撞击——这些宏大的事件对于像我这样的人而言太过抽象，甚至无法触发一丁点儿紧迫和重要的感觉，因此也比不上我左手小指上某根微不足道的小倒刺所引发的真实性的感觉。

我们都是以自我为中心的，对于我们而言最最真实之物，归根结底还是我们自己。在所有的事物中，最真实的是我的膝盖、我的鼻子、我的愤怒、我的饥饿、我的牙疼、我的侧腰疼、我的悲伤、我的欢乐、我对数学的爱、我的抽象天花板，不一而足。所有这些东西的共同点、亦是将它们联系在一起之物，正是由“我”的概念而来的“我的”这个概念。因此，虽然它不如一个鼻子乃至一次牙疼具体，但是“我”这个东西，在我们每个人的眼中，都形而上地构成了万物之中最为坚固且不容置疑的一块岩石。它有可能是一种错觉吗？或者，如果它不是一个纯粹的错觉，那么它有可能不如我们想象中那么真实和坚固吗？相比于一桶摸得着、举得起、挪得动的黄金，一个“我”有可能更接近于一道捉摸

不透、若隐若现、闪烁不定的彩虹吗？

运气不好，没成功，做不到

有一天，很多年前的一天，我想把装在书房地板上那个小纸箱里的信封都拽出来，叠成一组塞到我书桌的一个抽屉里。于是，我打开了纸箱，把手伸进去，用右手扣住里面那叠信封（大约有100封），然后用力紧压，好把它们一鼓作气地全部拽出箱子。这一切都没什么好惊讶的。但是突然之间，我在拇指和其他手指之间感受到了一样让我惊奇的东西。在那只摇摇欲坠的小小硬纸箱的正中心，竟然坐着（或浮着）一颗玻璃球！

同我们那一代的绝大多数美国人一样，我的手里拿过不下数百次玻璃球，可以毫无疑问地确定我摸到的那个东西是什么。亲爱的读者，我和你一样，曾经也是个“玻璃球老手”。但是一颗玻璃球到底是怎么混进我平时摆在桌子上的这个小盒子的呢？那时候我的孩子尚未出生，所以不可能是他们干的。而且不管怎么说，它怎么可能悬浮在盒子的正中心，而不是躺在盒子底部呢？重力的作用怎么失灵了？

我费力地在信封之间用眼睛寻找一颗光滑的彩色小玻璃球体。运气不佳，没有看到。然后我又把手指伸到信封的间隙去摸索翻寻，凭感觉去找它。还是没成功。但是再之后，当我又像刚才那样，把一整叠信封一把抓起的时候，它又出现了，还是那么实在！这个玻璃球小鬼到底藏到哪儿去了？

我找得更加仔细了，当然是把所有的信封都从盒子里拿了出来，然后试图把它从信封间抖落出来，可是仍然做不到。最后，我逐一检查了那些信封，发现每一封都空空如也。真该死，这到底是怎么一回事儿？

一首出其不意地献给老朋友“副”的颂歌

聪慧（当然，除此之外，还是一个信封老手）的读者啊，在你看来，事情大概已经很明显了，但是请相信我，我确实发懵了一两分钟。最终，我还是意识到，这里面本来就没有什么玻璃球，只有某种在那个玻璃球老手的感觉里同玻璃球一模一样的东西。它是一种副现象；而引发它的事实是，每个信封在其折叠出“V”字形的顶点处，都叠着三层纸和一层薄薄的胶。这一无心插柳的设计决策带来了一个意料之外的结果，当你挤压100张对齐叠在一起的信封时，你不能像挤压其他部位一样压紧那片小小的区域——它能抗压。你在手指上感受到的阻力硬度与一种我们更熟悉的（我敢说是“更真实”的吗？）硬度存在着一种诡异的相似性。

你大概还能从对前面章节中的回忆中想起，副现象是由微小的、通常是肉眼不可见或难以察觉的、极有可能是完全未知的大量事件所促生的一种集体性的、看似一致的产物。换句话说，一个副现象，就是由很多绝非错觉的小事件通过彼此之间的相互冲撞而创造出来的一个大规模的错觉。

好吧，我被盒子里的玻璃球这个副现象的错觉深深地吸引并迷住了，甚至还给这个装信封的小盒子起了个绰号，就叫“副”。从那会儿到现在，30年已经过去了，可我依然把它留在身边。（不幸的是，经过了这么长时间之后，那个盒子已经处在支离散架的边缘了。）有时候，当我出差去某地做关于“我”与自我概念的讲座时，我会随身带着副，并让现场的听众亲自伸手进来触摸并感觉，如此一来，副现象——在此即副现象——的概念，于他们而言就变得非常之真实和生动了。

就在最近，我还曾前往亚利桑那州的图森市，做了一场这样的讲座，也随身携带了副。现场的一位名为让奈尔·金（Jeannel King）的听众被副的传奇事迹深深打动，竟为它创作了一首诗，并用破格把它的经

历译入了她自己的人生。几天之后，她把诗发给了我。我反过来又被她的诗深深地打动，于是去征求她的同意，希望能把这首诗放在下面，而她也很慷慨地回答说，如果我真的这么做，她会十分高兴。所以闲话少说，下面请欣赏让奈尔·金在副的启发下所创作的一首悦人耳目的诗：

信封盒的颂歌

（献给每一位找不到玻璃球的人……）

让奈尔·金

地板上有一盒信封——

我想把它们收入抽屉。

我伸手捏住信封——有什么东西！

我向缝隙间看去——只有空气。

我又用手捏了捏，原来是颗玻璃球。

这颗玻璃球，是不是我心中的幻谬？

下定决心，一封一封翻找，

信封里面——还是什么都没找到！

为了仔细检查，我必须不顾纸刃划伤，

也不怕一粒粒的尘土飞扬。

第四十六张信封？还是在凌乱之中？

我的玻璃球整个儿消失无踪。

然后妈妈用粗哑的嗓音低声说，

“儿子，你再继续下去，会走火入魔！”

爸爸反对说：“你别多管闲事，

让儿子到信封里去寻找自己的上帝！”

于是，信封散落满地，

而我坐在地上，烦恼如斯。

我的玻璃球不见了，可在我的心里，

是不是也许还有更多的东西？

因为这个整体已经超越了部分：

没有任何部分可以独立存在。

拆分开来，玻璃球只是个玩笑。

为了完整，我要与盒子相守到老。

没有球面，没有半径，没有质量

关于我那颗作为副现象的玻璃球，其最为古怪的一个方面可能就是，我曾经那么确信盒子里的那个“东西”是球形的，那么自信满满地估计出它的直径长短（大约半英寸，和大多数的玻璃球一样），并且还能描绘出它有多么坚硬（通过比较的方式，比方说相对于一个鸡蛋黄或者一团黏土）。这个并不存在的事物的很多方面都是清晰而熟悉的有形现象。一言以蔽之，我是陷入了一个有形的错觉之中。那里根本没有玻璃球——只有一种统计学意义上的副现象。

尽管如此，依然不能否认，在向我的读者传达我的经验时，“它摸起来感觉就像是一颗玻璃球”这种说法，远比“我体验到了100层三重叠纸和100层胶水精准对齐而产生的集体效应”这种写法，更加透彻清晰。我仅仅是把它叫作了“玻璃球”，就能让你对我的感觉形成一个明确的印象。如果我没有使用过“玻璃球”这个词，你能推测出一叠厚厚的信封中央会引发出——简言之，这种集体效应感觉上像是一种非常简单和熟悉的物体——某种感觉起来像是完美的球形、似乎拥有尺寸且极为实在的东西吗？我对此深表怀疑。由此可见，虽然盒子里没有真实的玻璃球，但是保留“玻璃球”这个说法还是有意义的。这里面存在着某种感觉起来十分像是一颗玻璃球的东西，而这一事实对于我的描述以及你对于此中情况的把握都是至关重要的，这正如“走廊”、“星系”和“黑洞”这些概念对于我观察及描述那台观看自身的电视机屏幕上发生的现象是至关重要的一样——即便严格来说，在那个屏幕上并看不到任何走廊、星系或者黑洞。

责任止于此

我重新阐述了信封盒子里那个半真半假的玻璃球的故事，意在为某种真实性的类型提供一个喻体。当我们不可抗拒地感觉到某种“实在”且“真实”的东西居于我们自己的内核之中，便是应用了这种现实性，而也正是这种强烈的感觉，才使得“我”这个代词对于我们的存在变

得不可或缺且至为重要。这本书的论题就是，在一个脱离了胚胎和婴儿期的人类的头脑中，有一类特别的抽象结构或模式[注]，扮演着与纸张和胶水的精准对齐相同的角色——这种抽象模式引发了某种*感觉上*像是一个自己之物。我打算详述这种抽象模式的本性，但是在此之前，我不得不首先说明，“一个自己”这个说法是什么意思，或者更为具体地说，为什么我们看起来需要这样的一个概念。

每个活着的生命，不管其多么简单低等，都拥有一套内在的目标嵌于其中，这得益于在时间的长河中进化而成的、塑造了其所属物种的那个反馈环。这些反馈环都是些为人熟知的、几乎已是陈年滥调的生命行为，诸如觅食、趋温、求偶等等。有些生物还额外发展出了属于他们自己的个体目标，如演奏某些特定的音乐、参观某些特定的博物馆或者拥有某种特定类型的汽车。不管一个生物的目标是什么，我们都习惯说它在*追逐*这些目标，以及——至少当目标足够复杂时——我们还经常补充说，它这么做是因为它*想要*某些东西。

“你为什么骑车去那栋大厦呢？”“因为我想去那里练习钢琴。”“你为什么想要练习钢琴呢？”“因为我想要学会巴赫的那首曲子。”“你为什么想要学会那首曲子呢？”“我不知道，我就想学——它太优美了。”“但是这首曲子有什么特别的优美之处呢？”“我说不准——它就是以某种特别的方式打动了我。”

生物把它的行为归因于被它指称为*它的欲望*或*它的需求*之物，但是这并不能说清楚为什么它拥有这些欲望。在到达某一点后，便不再有更多分析或说明的可能性了；这些欲望就在那里，仅此而已，而对于生物而言，它们看起来就是它的决定、行为和动作的根本缘由。而且在表达它做什么以及为什么做的句子里，总是有“我”这个代词（或者它的姐妹词“自我”、“我的”等等）存在。看起来责任只能推至此处了——停在了所谓的“我”之上。

反诸自身的原动力

夏天刚刚与我挥别，带走了潮湿与闷热。秋日的午后，天气清爽怡人，我决心出门痛快地跑上一圈。我走进我的卧室，四处搜罗我的跑步短裤、跑鞋和T恤衫，迫不及待地换上行装，一转眼的工夫，我的身体已经发现自己来到了室外的人行道上，我的双脚轮流地蹬踏着地面，我的心脏开始扑通扑通地加快了跳动。还没等我回过神来，我已经跑出了100步，不久之后，就达到了300步。然后是1 000步，接下来是3 000步，我仍在全速前进，气喘吁吁，大汗淋漓，同时在心里对自己说，"我为什么总是告诉我自己，我喜欢跑步呢？我恨它！"可是我的身体并没有停歇片刻，而且不管我的肌肉有多么疲累，我的自己只会像一名有虐待癖的中士教官训讽一批新兵那样对它们说，"别想当逃兵！"——结果你瞧，我那可怜兮兮的、上气不接下气的、不听使唤的身体，竟无条件地服从了我的自己，甚至违逆它本身的意愿而加速冲上了陡峭的山坡。我那不着形迹的"我"抱着同样无影无形的在这个秋日跑步的决心，相当无情地使唤着我那具穿着运动短裤的叛逆的身体。

所以，是谁在此使唤着谁？在这幅图景里，指使我们行动的物理粒子在哪里？它们是看不见的，而即便你记得它们的存在，它们看起来也只不过是次要的角色而已。是那个"我"，即那个贯通了欲望与信念的集合，发动了这里的每一件事情。那个"我"才是发动了所有生物行为的原动力，是那个隐居幕后的神秘推手。如果我想要某件事发生，我就会用意志力驱使它发生，而但凡它没有超脱我的掌控范围，它就通常会真的发生。构成身体的分子，不论其在手指、胳膊、大腿、喉咙、舌头或是任何部位，都会俯首帖耳地遵从那高高在上的伟大的"我"所发出的至高命令。

由此而言，我踩下汽车的脚踏板，然后自然而然，我那辆重达一吨的汽车便服帖地径直开往了我想要去的地方。使唤这个庞大物体的，正是那个缥缈无形的"我"。我拿起筷子，然后顺理成章，豆角听话地就到

了我的碗里，而我则收获了自己贪求的感官喜悦以及自己需要的营养。我在苹果电脑的键盘上打字，然后果不其然，句子顺从地出现在电脑屏幕上，基本上表达了那个缥缈的“我”希望表达的思想。而在所有这些情况之中，粒子们去哪儿了？根本看不见。看起来唯一可见的，就是那个令一切得以发生的“我”。

如此一来，如果是这个“我”引发了一个生物所做的每一件事，如果这个“我”可以对这个生物所有的决定、计划、行为和动作负责，那么这个“我”显然至少是存在的。

上帝之眼对闹壳之眼

说到这里，我想重新说回闹壳那个意象。在我关于闹壳里飞速移动的微型的小磁交弹和更加庞大也更加慵懒的小磁交弹球的讨论中，处于核心地位的一个事实是，这个系统可以在两个相距甚远的层级上分别被观察到，并由此引导出两种相差远甚的解读。

从更高层级的“思考动力学”视角出发，这里发生的是一种符号化的活动，小磁交弹球彼此之间交互作用，而它们所利用的“热能”则来自那些肉眼不可见的小磁交弹的翻滚蒸腾。从这个视角出发，则我们看到的每一个小磁交弹球事件都是由一系列其他的小磁交弹球事件所引起的，虽然其中因果关系的细节往往扑朔迷离或太过模糊，以至于难以精确把握。（我们对于这种日常生活中因果性的模糊面目并不陌生——例如，如果我在篮球场上非常可惜地投丢了一个罚球，那么我们都知道这是我的错，而且是我在某个环节上出了一点岔子，可是我们并不知道具体错在了哪里。如果我投出一枚色子，正面是“6”，我们丝毫也不会感到惊讶，可我们同样也不知道为什么投出的正面是“6”——我们根本不会思考这个问题。）

反之，从更低层级的“思维统计学”视角出发，则有且只有小磁交弹是真正存在之物，它们的交互作用是以基础性的彼此冲撞和击打的动态完成的——而且从这个视角出发，在因果性上便不再有任何模糊或疑问了，因为每件事都处在清晰、精准、棱角分明的数学法则的掌控之中。（如果我们能够任意放大我们的胳膊、手和手指以及篮球、篮板和篮筐或者色子和桌子，然后以我们想要的任何慢速观察每件事发生的慢动作，那么我们就能准确地发现到底是什么造成了罚球失误或者投到“6”。这可能需要一路降低到原子的层面，但是无所谓——最终，原因总会清晰地浮出水面。）

如果一个人对闹壳有充分的理解，那么在其看来，两种视角便都是合理有效的。只不过，后者可能看起来更加基础，不遗漏任何细节（我们可以称之为“上帝之眼”的视角），而前者则丢掉了大量信息，完成了高度浓缩的简化，因为它更有效率，所以看起来可能对于我们这些凡人而言也更加有用（尽管有些事情的发生看起来“毫无缘由”——可那正是视角转换的代价）。

我不是上帝

但是，并不是所有观察闹壳的人都享有这种奢侈的能力，可以在两种天壤之别的视角间来回跳换。也并不是所有会思考的生物都能像我在第3章中描述的那样充分或清晰地理解闹壳。很简单，上帝之眼的视角并不会开放给所有的观察者；事实上，某些观察闹壳的人可能压根儿没有想到过还有这种视角的存在。而我在心中特别想到的是一位身份特殊而享有先天优势的闹壳观察者，那就是闹壳本身。

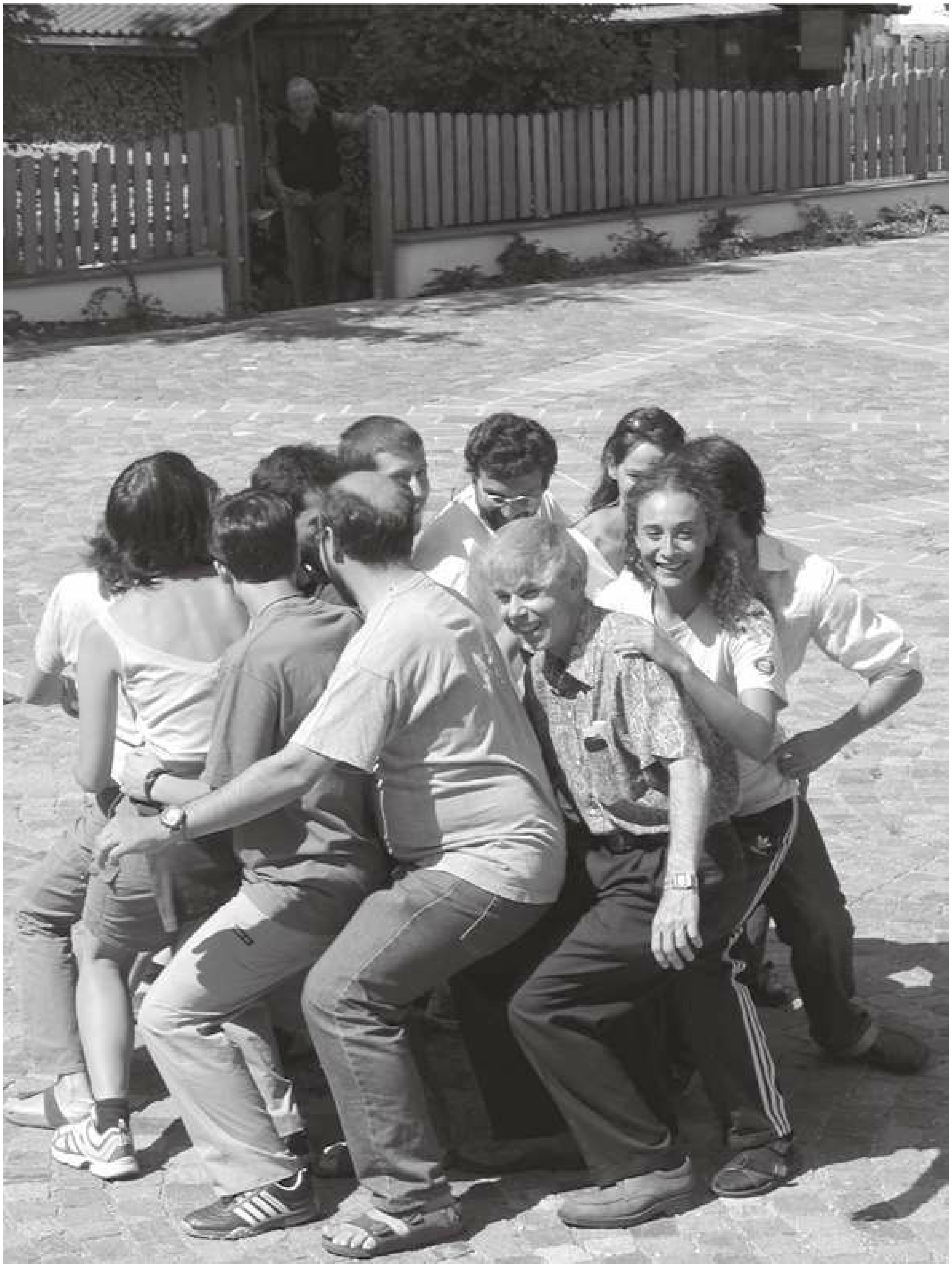
当闹壳费心琢磨它自己本性的时候，特别是当它正在“成长”的过程中，刚刚开始认识自己，并且远在成为一名研究数学和物理（以及最终可能成为闹壳学）的科学家之前时，它所能意识到的就只有它的小磁交

弹球行为，而看不到它在小磁交弹层面上的翻腾。毕竟，正如你我所知（而它并不知道），闹壳对于所有事物的感知都是极其粗线条的简化状态（被大批冲击信号的风暴所集体触发的一小组小磁交弹球）——而它对于自己的感知也不例外。

天真无知的闹壳对于深藏幕后的一切毫无所知，不知道在某个看不见的微观层面上，发生着翻滚、蒸腾、小磁交弹层面的行为。哪怕只是在理论的层面上，它也从来没有想到过还存在着任何其他观察其本性与行为的视角。没错，这个年轻的闹壳让我想起了青少年时代的自己，那时我还没有读过菲佛（Pfeiffer）那本以及彭菲尔德（Penfield）和罗伯茨（Roberts）合著的那本讨论人类头脑的书[注]；那些书曾令我心神不宁，却也同样点燃了我的想象力。这个理想主义的年轻闹壳很像那个天真无邪的10多岁的少年道格，他刚刚来到那个转折的关口，马上就要开始投入到每一个人类闹壳内部的那片无缝的黑暗当中，瞥见每日每夜发生在那里的诡谲之事。

因此，在闹壳对自己的前科学认识中，深深地嵌入了一种把自己视为完全由思想和观念驱动的生物，这种认识像黄岗岩和大理石一般坚硬不移。它的自我意象，与一个命运完全由数十亿个肉眼难见的冲击相撞的微观事物所主宰的巨大机械化实体可谓是大相径庭。相反，这个天真的闹壳镇静自若地宣明自身："我只受我自己的驱使，而非其他地方的任何事物。"

那么，这个"我"到底是一种什么东西呢？闹壳认定是它驱动了自己的选择和行动，而人类也同样如此认为。如果现在我说，"我"是一种抽象的锁定在大脑或闹壳之内的奇异的环路——事实上，就是一个怪圈——想必已经不会令人感到惊讶了。因此，为了清晰地阐明为我所称的构成"我"之物，我有必要详细说说我口中的"怪圈"是什么意思了。而且，既然我们马上就要结束《我是个怪圈》的第7章，差不多也到时候了！



1. 但凡有人愿意尝试，就请试着把*Ash and Gross*这本书从头读到尾吧。这有点像在一家

地道的印度餐馆点了一道“印度辣”的菜——你会怀疑自己为什么要那么做。

2. The great Galois was indeed a young radical, which led to his absurdly tragic death in a duel on his twenty-first birthday, but the phrase “solution by radicals” really refers to the taking of *n*th roots, called “radicals”. For a shallow, a medium, and a deep dip into Galois’ immortal, radical insights into hidden mathematical structures, see [Livio], [Bewersdorff], and [Stewart], respectively.

3. “Real Patterns” in [Dennett 1998] argues powerfully for the reality of abstract patterns, based on John Conway’s cellular automaton known as the “Game of Life”. The Game of Life itself is presented ideally in [Gardner], and its relevance to biological life is spelled out in [Poundstone].

4. 菲佛的书是指*The Human Brain*(1961)；彭菲尔德和罗伯茨合著的那本书是*Speech and Brain-Mechanisms* (1959)。——译者注

第8章
开启一段怪圈之旅

折盖环路模式，大腿环路模式

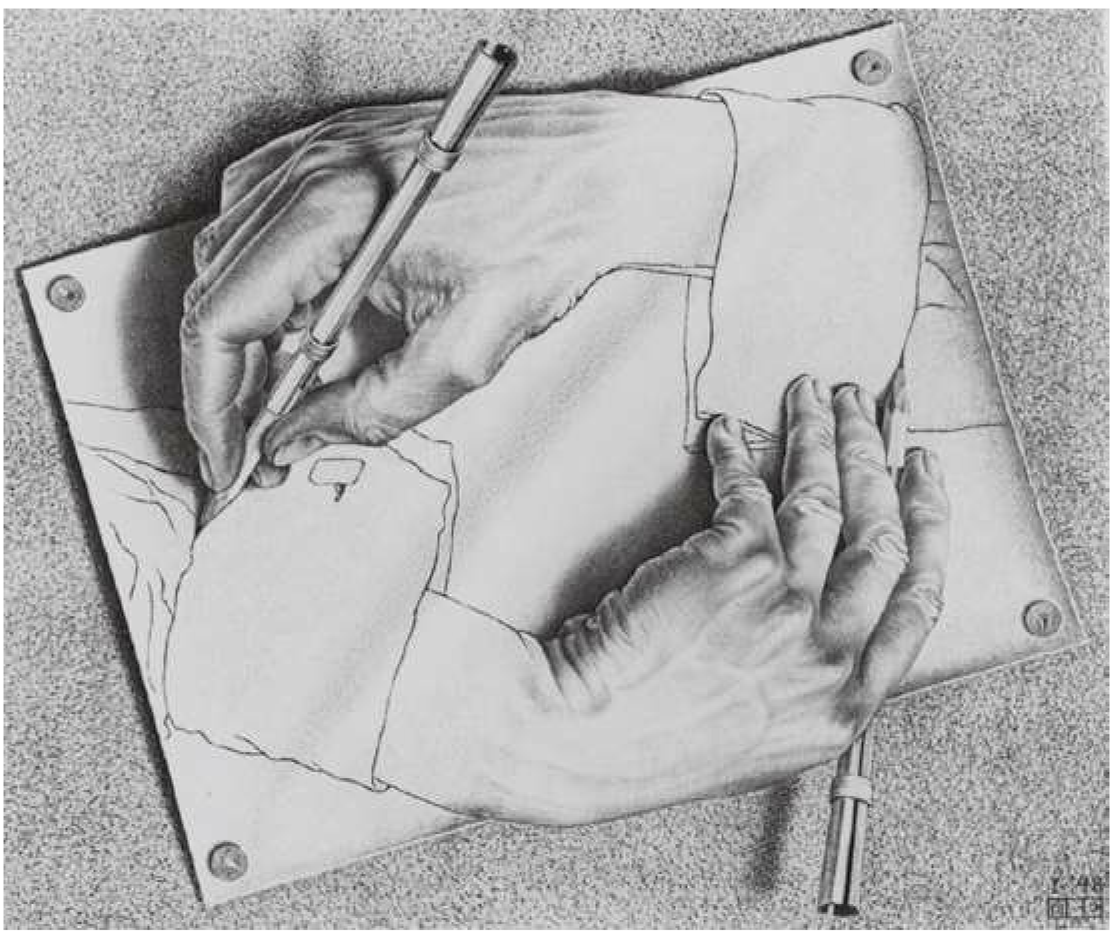
我在第4章中已经描述过，自己在小时候曾多么迷恋通过把纸板箱的四片折盖按照一种轮转的顺序互相折叠起来而合上箱子的厚颜行径。完成那最后一下禁忌的折叠，总能给我带来一阵战栗的快感（甚至直到今日，这种快感仍未完全消失），感觉自己正在同悖论性展开一场危险的挑逗游戏。然而不消说的是，真正的悖论却从未到手。

这种“折盖环路模式”还有一个近亲，就是前一页图中所示的“大腿环路模式”。在这张摄于安泰塞尔瓦·迪·梅佐的照片里，我位于前排中央，正在咧嘴大笑（我在此称自己为“A”），并坐在一位年轻女子（“B”）的大腿上，B也在咧着嘴笑，并坐在C的大腿上，而C又坐在D的大腿上，如此类推，直到K坐在我的大腿上，形成了一个完整的环路。一条大腿上承载着很多条大腿，却并没有被压垮。如果你从没玩过这个游戏，我建议你去试一试。到底是什么东西在支撑着这个环路不被压塌，是个相当令人困惑的问题。

同折盖环路模式一样，大腿环路模式也与悖论打了一个擦边球。因为虽然环路中的11条大腿都在用力支撑着另一条大腿，但很明显的是，这样一个大腿环路模式是可以在物理世界中实现的，所以它不是一个真正的悖论。即便如此，当我在这个大腿环路模式中扮演“A”的角色时，我觉得仿佛是自己——虽然是间接地——坐在了自己的大腿上！这是一种极其奇怪的感觉。

在艾舍尔的画中寻求怪圈

尽管如此，当我说到“怪圈”的时候，我心里想的却是另外一回事——那是一种没那么具体的、更难以捉摸的概念。我所说的“怪圈”，指的是——不管怎么样，先在这里给出一次最初的尝试——一种抽象的环路，而并非一种物理的回路。在一系列构成环路的不同阶段中，从一个抽象（或结构）层级向另一个层级的转移，感觉像是沿着一种等级结构在向上运动，结果这种连续不断的“向上”转移，却以某种方式形成了一个闭合的环路。也就是说，虽然感觉上是离开得越来越远，结果最终却令人惊讶地回到了出发的原点。简言之，一个怪圈就是一个悖论式的层级交叉的反馈环。



艾舍尔的石版画《画手》（*Drawing Hands*，见上图）是最经典

（我要很遗憾地说，如今已经不再新鲜[注]）的一个例子。在这幅画里，可以看到（取决于一个人从哪里开始观看）一只右手正在画一只左手（到目前为止，还没有出现什么悖论式的东西），结果那只左手却也正在画那只右手（突然之间，出现了一个深度的悖论）。

在这里，层级上的抽象转换是从被画者到画者（或者同样可以说，是从图像到画家）之间的向上跃迁，后者的层级在直觉上"高于"前者，而其理由也不止一二。首先，画者总是一种有灵的、动态的生命，而被画者则是一副定格的、不动的图像（画的内容可能是一个非生命的物体，也可能是一个有生命的实体，但不管怎样，都是静止不动的）。其次，画者是三维的，而被画者是二维的。最后，画者选择画的内容，而被画者对此没有任何发言权。于是，至少在这三种意味上，从被画者到画者的跃迁总会带有一种"向上"的感觉。

正如我们刚才所说，从任何一幅被画的图像到它的画者，在本质上都存在着一种鲜明而清晰的向上跃迁——可是在《画手》里，这种向上的法则遭到了鲜明而清晰的违反，因为每一只手都在等级上"高于"另一只！这怎么可能呢？好吧，答案再明显不过了：这整个就是一幅被画出来的图像而已，仅仅是一种幻想。但是因为它看上去如此真实，因为它如此有效地把我们带入了它的悖论式的世界之中，所以它至少在很短的时间里使我们相信了它的真实性。不仅如此，我们也很乐意接受这样的欺骗，所以这幅画才会如此大受欢迎。

《画手》中的抽象结构本可以为真正的怪圈提供一个完美的范例，只可惜它还有那样一个小小的瑕疵——我们以为自己看到的东西，并不是真的；它是假的！不可否认，它的画工是如此天衣无缝，让我们似乎看到了一个成熟、忠实、正经的悖论——但我们之所以产生这种信念，只是因为我们把自己的怀疑悬置了起来，在思维上滑入了艾舍尔那个充满诱惑的世界。至少在短暂的时间里，我们陷入了一种错觉之中。

于反馈中寻找怪圈

那么说到底，真正的怪圈——毋庸置疑地属于我们所生活的世界却具有悖论性的一种结构(注)——究竟是否存在呢？还是说，所谓的怪圈只不过是与悖论打擦边球的错觉，是与悖论隔空挑逗的幻想，像瑰丽迷人的肥皂泡，一触即破呢？

视频反馈怎么样？我们的这位老朋友，能不能成为这种怪圈的候选人呢？遗憾的是，虽然这种时髦的现象具有循环的特点，而且游走在无限性的边缘，但是它一点儿都不具有悖论性——同它那位更简单、更古老的近亲，音频反馈，没什么差别。可以肯定的是，把电视摄像头指向屏幕（或者把麦克风摆到扩音器的跟前）会让人产生一种奇怪的玩火的感觉，不仅因为这违反了一种看似自然的秩序，而且仿佛创造了一种真正无限的回归——但是我们只要用心想一想，就会意识到，其实从一开始就不存在什么严格的等级秩序，而假想中的无限性也从未抵达过；然后，肥皂泡便被戳破了。所以，虽然这种类型的反馈环无可置疑地形成了环路，而且令人还感觉有一点奇怪，但是它们并不是“怪圈”这一范畴内的成员。

在罗素的阴郁中探求怪圈

幸运的是，并非错觉的怪圈确实存在。我之所以要说“幸运”，是因为这本书的命题便是，我们自己——不是我们的身体，而是我们的自己——就是怪圈，所以如果所有的怪圈都是错觉的话，我们便也都是错觉，那可就太不妙了。所以，在现实世界中存在某些怪圈，自然是一件相当幸运的事。

另一方面，呈现出一个让所有人都看得见的怪圈，绝非易事。怪圈是害羞的生物，不愿意在日光下见人。事实上，这种现象的典型事例直

到1930年才被库尔特·哥德尔首次发现，而它在被哥德尔发现之前的藏身之处，不在别处，恰恰就在罗素搭建的那座阴郁、清冷、据称可抵御一切悖论的类型论的城堡之中。

一名24岁的奥地利逻辑学家，到这座令人望而生畏的冷峻的英国堡垒中四下窥探，用意何在？他就是对悖论太着迷了而已。虽然他知道，罗素和怀特海本应已经把所有的悖论驱逐殆尽了，可是他的直觉仍然告诉他，在极端丰富并充满弹性的数字的本性之中，还深藏着某种滋生悖论的东西，即便是最为干旱贫瘠的沙漠或者寸草不生的花岗岩宫殿，也无法阻止悖论的生长。哥德尔的这种怀疑，是由当时接二连三出现的悖论所唤起的，这些悖论在与数字打交道时，使用了种种奇妙的新方式，而虽然有些人声称自己有办法将它们一一化解，哥德尔却在心中深信，这些智力游戏的背后，存在某种更加深刻的东西，

牛津大学图书馆的贝里先生

在这些光怪陆离的悖论中，有一则是由一位名叫G. G.贝里的牛津大学图书馆管理员(注)于1904年提出的，当时距哥德尔出生还有两年。贝里对于用文字描述数字的微妙可能性很感兴趣。他注意到，只要你足够用心，就可以为几乎每一个整数找到一则相当简约的描述。比如，整数12只需要一个音节（twelve）为其命名，而整数153只需要用四个音节便可以精确地描述（twelve squared plus nine或者nine seventeens），整数1 000 011的命名只需要六个音节（one million eleven），等等。那么，你最少要用几个音节才能描述出数字1 737呢？

通常而言，人们会认为，数字越大，对于它的描述就会越长，但是这都取决于用“界标”整数——少数名字或描述极短的整数——表达这个数字的难易程度，比如10的一万亿次方（ten to the trillion）这个数字，只需要相当节约的五个音节便可描述。当然，大多较大的数字，既不是

界标整数，离界标整数也不近。确实，到目前为止，绝大多数的数字都是"无名的"，只能用一串又长又复杂的描述来表达，因为它们就是"很难描述"的。它们就像设立在偏远地区之外很遥远的前哨基地，想要抵达那里，只能经由一系列漫长的小路，而当距目的地越来越近的时候，这些小路只会变得越来越窄，越来越颠簸崎岖。

想一想777 777这个数字。它的标准英文命名是"seven hundred seventy-seven thousand seven hundred seventy-seven"，足足有20个音节之多，可谓相当之长。但是这个数字还有另一种相对较短的描述方式："seven hundred seventy-seven times one thousand and one"（"777乘1 001"），只有15个音节长。节约了不少！而我们还可以把它进一步压缩成："three to the sixth plus forty-eight, all times ten cubed plus one"（3的6次方加48，再乘以10的3次方加1）或者直白的"the number whose numeral is six sevens in a row"（由连续6个7组成的数字）。这两种表达方式，都只需要14个音节而已。

只要足够用心，我们可以为777 777这个数值找到许多英语表达方式，而按照其中的某些方式把它读出来，可能只包含很少的音节。比如，"seven thousand seven times one hundred eleven"（"7 007乘111"）怎么样？只剩下13个音节了！而"nine cubed plus forty-eight, all times ten cubed plus one"（9的3次方加48，再乘以10的3次方加1）呢？减少到12个音节了！"thrice thirty-nine times seven thousand seven"（三个39乘以707）呢？减少到11个音节了！我们还能把这个数字继续压缩到多短呢？这可不容易看出来，因为777 777完全有可能拥有某些精妙的数字性质，让它可以用十分简约的形式表达出来。这样一种描述甚至可能需要引用到比777 777这个数字本身大上很多的界标整数。

图书管理员贝里在反复思索了这种寻找更短描述的动作的微妙本性之后，恶毒地提出了一个性质十分特别的数字，我为了纪念他，把这个数字命名为b：b是至少用30个音节才能在英语中得到准确描述的最小整

数。换句话说，能够精确表达*b*的数值的描述，绝不会少于30个音节。既然描述这个数字总是需要如此多的音节，可知*b*是一个很大的整数。只不过，大约而言，*b*到底有多大呢？

凡是你在一张报纸、一份杂志或者一本天文学或物理学教科书上看到的数字，几乎可以肯定不会超过10多个音节，最多20个音节也到头了。例如，阿伏伽德罗数（6×10^{23}）可以用一种十分简化的方式给定（“six times ten to the twenty-third”，6乘以10的23次方——仅仅八个音节而已）。要想找到一个不管怎么描述它都至少要用30个音节的那么大的数字，绝非易事。

不论如何，根据定义，贝里的*b*是不能在我们的口语中被浓缩到30个音节以下的第一个整数。我再重复一遍，并用着重号加以强调：它是在英语的描述中永远要用到至少30个音节的最小整数。但是稍等片刻！我用着重号呈现的短语包括了多少个音节呢？数一数—24个。我们莫名其妙地用少于它的定义所规定的最小数量的音节，把*b*描述出来了。事实上，加着重号的短语并不仅仅是“莫名其妙”地描述了*b*：它正是*b*的定义！所以，*b*的概念恶毒地挖空了自己的墙角。有某种非常奇怪的事情发生了。

我不能告诉你它有多么难以形容！

巧合的是，在日常语言中还有一些很常见的说法，也具有类似风格的自我削弱的品质。以“难以形容”这个形容词为例，如果我说，“他们的房子真是难以形容”，你必然会从我的话里得到某种视觉意象——尽管（或者不如说恰恰因为）我的形容词表达的意思是，没有与它相符的描述。甚至还有更加古怪的说法，如“那辆卡车的轮子真是大得难以形容”或者“我对于你的感激难以言表”。然而，这种自我削弱的品质对于交流却是至关重要的。

贝里的悖论还有一种“初级版本”，是在这一悖论提出几十年后出现的，如下所述。某些整数是有趣的。0是有趣的，是因为0乘以任何数字都等于0。1是有趣的，因为1乘以任何数字都等于那个数字。2是有趣的，因为它是最小的偶数。3也是有趣的，因为它是最简单的二维多边形（三角形）的边数。4是有趣的，因为它是第一个合数。5是有趣的，因为它是三维正多面体的数量（这只是诸多原因之一）[注]。6是有趣的，因为它不仅有3个因子（3×2×1），还是3个因子相加之和（3+2+1）[注]。我可以如此无限列举下去，但我想你已经明白了。问题在于，我们什么时候才会遇到第一个无趣的数字？可能是62？还是1 729呢？好吧，不管是哪个数字。作为第一个无趣的数字，对于任何数字而言都是一个有趣的性质！所以62（或者不论你选择哪个数字作为候选）结果还是会变得有趣——它有趣，正是因为它无趣。如此一来，“最小的无趣数字”这一观念就反噬了自身，显然与贝里*b*的定义反噬自身的方式如出一辙。

如我们所知，正是这种语言的反转触痛了罗素那颗敏感脆弱的内心，而且也正是罗素本人第一次公开了贝里的悖论数字*b*——这也算是他的功劳了。1906（四个音节！）是哥德尔出生的年份，而罗素也在这一年讨论了这个悖论。他在自己的文章中，绞尽脑汁地想要化解由这一悖论带来的痛楚，声称它不过是一种错觉，由在数学语境中对英语单词“describable”（可描述的）愚昧的误用所引起。罗素称，这个概念必须放置在由可描述性的不同类型组成的无限等级序列之中——在第0等级上的描述，只能征用纯粹的数字；在第1等级上的描述，可以使用算术，但是只能征用在第0等级上的描述；在第2等级上的描述，也可以使用算数，但是也只能征用在第0和第1等级上的描述；如此依次类推。由此，罗素自信已发现了新的真理，并指出，“可描述性”这个观念，如果不加以某种特定等级的限制，就只不过是一种胡想罢了。罗素以这种全新的理论（全新的类型论）为武器，宣布自己已经把那个珍贵而精致的严谨推理的世界保护起来，令其不会再受到来自“贝里果”（Berry-

Berry）那种丑陋而恶心的瘟疫的困扰。

模糊性将贝里埋葬

虽然我与和罗素一样，也觉得在贝里的悖论中有点不太对劲儿的地方，但是关于这个不对劲的地方到底在哪儿，我和他的看法却大相径庭。我关注的不妥之处是，在对于数学陈述的表达上，英语是一个模糊到无可救药的语言；它的词汇和短语都太模糊不清了。第一眼看上去似乎是精确的说法，结果都会引入大量的歧义。例如，我在前面曾用“nine cubed plus forty-eight, all times ten cubed plus one”（9的立方加48，再乘以10的立方加1）这个说法来描述777 777，但它其实是有歧义的——比如，它可以被解读为表示777乘以1 000，最后加上1，于是得到的结果便会是777 001了。

而这种小小的歧义性只是冰山一角。事实的真相是，到底哪种英语表达可以算作对一个数字的描述，本身就是说不清楚的。看看下面这些短语吧，它们都意在描述某个特定的数字：

- the number of distinct languages ever spoken on earth（在地球上被说过的不同语言的数量）
- the number of heavenly bodies in the Solar System（太阳系中的天体数量）
- the number of distinct four-by-four magic squares（不同的4乘4魔方的数量）
- the number of interesting integers less than 100（小于100的有趣整数的数量）

它们都有什么毛病呢？好吧，它们都包含了定义不清的概念。

比如，一种“语言”意味着什么？手语算是一门语言吗？它是“被说”的吗？在语言和方言之间，有截然分明的界限吗？从拉丁语到意大利语，之间相差了多少种“不同的语言”？从尼安德特人时代到拉丁时代，人们说过多少种“不同的语言”？教廷拉丁文是一种语言吗？儿童黑话(注)呢？即便我们把人类在过去几百年中在地球上说过的每一句话都录了下来，想要客观地把每一句话归类到某种特定的“官方”语言，然后清楚地梳理开每一种“真正不同”的语言，最终搞清楚它们的数量，这种想法也仍旧是痴人说梦。就算只是说，要数清一间车库里的所有“物品”，都已经相当不切实际了，更别提有史以来的所有语言了！

再往下说，什么才算是一个“天体”呢？人造卫星也算吗？宇航员随机丢在宇宙中的漂流物和废弃物也算吗？我们要算上每一颗小行星吗？漂浮在土星行星环里的每颗不同的石头算不算呢？尘粒呢？飘散在真空中的独立原子呢？太阳星的范围有多大呢？此类问题，无穷无尽。

你可能会反对说，“但是，那些都不是数学概念啊！贝里的观念是要使用整数的数学定义。”说的没错，那么请给我在数学与世界的其他部分之间画一条截然分明的分割线吧。例如，贝里的定义使用了“音节数”这个模糊的概念。在finally、family、rhythm、lyre、hour、owl这些单词中，各有多少个音节呢？但是没关系，假设我们已经建立起了一套严格而客观的数音节的方法。可是，问题还在，到底什么才算是一个“数学概念”呢？数学这个学科本身真的有一个明确的定义吗？比如，“魔方”这个概念的精确定义是什么呢？不同的作者赋予它的定义都不相同。我们有必要在数学群体中开展一场民意调查吗？如果有必要的话，那么哪些人才算是这个模糊的群体中的成员呢？

“有趣数字”这个概念又如何？我们能给它提供某种数学上的精确性吗？正如你在前文所见，称一个数字“有趣”的理由可能包括几何以及数学的其他领域——可还是那个问题，数学的边界在何处？博弈论是数学的一部分吗？医学统计呢？植物卷须的扭转理论呢？如此等等，不一而

足。

总的来说，“整数的英语定义”这个概念本来就是一个无望的泥潭，因此，贝里提出的*b*的扭曲概念，同艾舍尔笔下两只互相画的手的扭曲概念没什么不一样，都是充满想象力的天才的虚构，而并不是一个真正的怪圈。又一位大有前景的怪圈候选人掉队了！

虽然在上面这段简短的小插曲中，贝里在1904年产生的想法看起来好像十分天真，但我必须指出的是，大约过了60年以后，年轻的数学家格里格·蔡廷（Greg Chaitin）在贝里这一想法的启发下，构想出一个更加精准的版本，不再使用英语，而是用计算机程序进行描述，而这次机智的转换，为哥德尔在1931年提出的定理提供了焕然一新的证据与视角。自此出发，蔡廷与其他学者进而发展出[注]一条重要的数学新分支，即著名的“算法信息理论”。再往下说，我们就离题太远了；我只是希望传达出贝里的奇思妙想中所蕴含的丰富性，因为正是它为哥德尔的革命性观念提供了滋生的土壤。

花生酱和浆果三明治

罗素为了杜绝贝里的悖论式建构，设立了一种杜绝所有自指性语言表达或自我包含的集合的形式主义，这不仅失之草率，而且谬之千里。为什么这么说呢？最近，我的一位朋友告诉我，她的一位朋友，一位年轻而天真的母亲，也设立了一条罗素式的禁令。那位女士以一种满怀善意的姿态，严格禁止任何玩具枪出现在她的家里。这条禁令发挥了一段时间的作用，直到有一天，当她给自己上幼儿园的儿子做好了一个花生酱三明治后，小伙子迅速地把三明治咬成一个手枪的形状，然后把它举起来，指着她大叫：“砰！砰！你死了，妈妈！”这则充满反讽意味的逸事阐明了一个重要的道理：不论你的禁令有多么严格，未禁之物很可能最终变成可塑性很强的中介物，并刚好塑造出你所禁止的那个东西。而

事实上，罗素对于贝里的驳斥几乎没有起到任何效用，因为在迈向20世纪的那个世纪之交，在那个知识暴动的年代，越来越多的悖论被创造（或发现）出来。坊间流传着这样一种说法：各种各样古老悖论的现代变体，在逻辑严密的数字世界中纷纷冒出来，而这个世界原本是一片天堂净土，没人曾想象过能有悖论生于此间，也从未有人在这个世界里见过这类东西，于是当悖论出现时，就会发生一些奇怪的事。

虽然这些新式的悖论给人的感觉像是在破坏那美好而神圣的推理和数字世界（或者不妨说，正是因为这一令人担忧的事实），相当多的数学家都大胆地踏上了一条寻找更深度、更麻烦的悖论的征程——也就是说，他们在探寻自己的学科根基有史以来面临的最有力的威胁！这听起来颇有点违反常理，但是他们相信，从长远的角度来看，这样一番求索对于数学是非常有益的，因为它会揭示出数学最为关键的弱点，展现出根基的动摇之处，以便有的放矢地进行加固，令其更加坚不可摧。简而言之，深深地扎入悖论的新潮之中，似乎对于任何一个研究数学基础的人而言，即便并非必不可少，也是一种有用的行为，因为新的悖论打开了关于推理本性——于是也关乎思维的难以捉摸的本性——的深刻问题，因而也事关人类思维本身的神秘本性。

一名为自己写自传的狂妄小生

我在第4章中曾提过，14岁的时候，我偶遇了欧内斯特·内格尔和詹姆斯·纽曼的小册子《哥德尔证明》，而且通过这本书，我对于循环悖论的观念着了魔，而那正是哥德尔的课题所围绕的中心。与我人生的那段时期联系在一起的一个更奇怪的怪圈，就是我恰恰在那个时候结识了内格尔一家。他们本来住在曼哈顿，但是在1959—1960学年“西下”斯坦福，而鉴于内格尔和我的父亲是老朋友，我很快就认识了他们全家人。内格尔家离开斯坦福之后不久，我就拜访了他们坐落在佛蒙特州布拉特尔伯勒附近那片平缓山区内的避暑别墅，在绿影青葱的后院里，我为他

们家的大儿子、我的朋友桑迪（Sandy）大声朗读了《哥德尔证明》的整本书，品味着一种扭曲的快感。桑迪和我同龄，而且我们都在以一种只有青少年才懂的狂热的痴醉感在探索数学的奥秘。

在如此强烈地吸引着我的力量之中，有一部分是来自哥德尔课题中心的奇怪的循环。但是我强烈的好奇心的另一半，在于我感受到的哥德尔在真正探索的东西，也是很多受到他启发的人所真正探索的东西，那便是人类思考机制和人类思维的秘密。哥德尔那篇1931年的文章一鸣惊人地亮出了那么多的问题——这些问题包括：

> 数学家在完成他们最有创造力的工作时，他们的脑中都发生了什么？
>
> 难道总是循规蹈矩的符号运算，并由一套固定的公理推导出定理吗？
>
> 一般而言，人类思考的本性是什么？
>
> 在我们脑中发生的事情只是一种决定性的物理过程吗？
>
> 果真如此，严格的法则管控着构成我们大脑的看不见的粒子，那么所有人，不管有多少奇特和闪光之处，都只不过是这些法则的奴隶而已吗？
>
> 从一套管控微观事物或数字模式的严格法则之中，能诞生出创造性吗？一台受法则约束的机器，能具有跟人类一样的创造力吗？
>
> 一台编好程序的机器，能够形成事先没有编入程序的观念吗？一台机器能做出属于它自己的决定吗？它有属于自己的观点吗？会迷惑吗？会知道它自己的迷惑吗？会不确定它是不是在迷惑吗？会相信自己有自由意志吗？会相信自己没有自由意志吗？会有意识吗？会怀疑自己有意识吗？它拥有一个自我、一个灵魂、一个“我”吗？会相信它对于“我”的狂热信念只是一个错觉，但又是一个无法避免的错觉吗？

关于元数学的理想主义之梦

回想青年时代的那些任性时光，每当我走进一家大学书店时（而我只要能多去，就不会少去），我会立刻窜到数学书区，搜刮每一本关于符号逻辑或者符号和意义的本性的书。就这样，我买了一本又一本有关这些话题的著作，比如鲁道夫·卡尔纳普（Rudolf Carnap）那本声名昭著又骇人听闻的《语言的逻辑句法》（*The Logical Syntax of Language*）和理查德·马丁（Richard Martin）的《真理与外延》（*Truth and Denotation*）。不用说，还有数不清的符号逻辑学教科书。我仔细地阅读过几本这样的教科书，但是卡尔纳普和马丁的巨著则一直静静地坐在我的书架上，一边威慑一边向我挑衅，总是看起来离我仅一步之遥。它们厚重而繁密，几乎无孔可入——但是我不停地在想，但愿某一天，在某个伟大的日子里，我终于能够阅读它们，完全地理解它们，然后最终穿透它们，抵达思维、意义、创造性和意识之谜的核心。如今回头再看，那听起来简直幼稚到离谱（首先是想象这是一个可以达到的目标，然后又以为所有的秘密刚好都藏在那些书里），但在当时，我真是一个虔诚的信仰者！

我在16岁的时候，有过一段不同寻常的经历，就是在斯坦福小学（我自己的小学母校）上符号逻辑课。我使用的是一套全新的教材，作者是哲学家和教育家帕特里克·苏佩斯。他正好住在我家所在的那条街道南边，而且他那本经典的《逻辑导论》（*Introduction to Logic*）也曾是我最依赖的指导书之一。苏佩斯正在进行一项实验，想看看严密的逻辑推理模式能否像算术一样灌输给孩子。学校的校长在我上学时就已经认识了我，有一天，他在学校的大厅里偶然撞见了我，便问我是否愿意给六年级班（我妹妹劳拉也在班上）开一门符号逻辑课，每周三次，上满一年。我欣然接受了这个提议，并且在接下来的一整年里，都彻底专心于此，乐此不疲，尽管也有几个孩子时不时地给我找点麻烦（比如拿橡皮筋打眼睛之类）。我教会我的学生使用很多推理规则，包括悦耳动

听的modus tollendo tollens（假言推理）和令人耳目一新的“假言三段论”，而与此同时，我也在打磨自己的技术，不仅是作为一个新手逻辑学家，也是一名老师。

驱动这一切的——我内在的热情——是我想要看到人类思维状态之谜揭开面纱的熊熊欲望。我渴望借此理解，在人类的大脑里，每分每秒都在发生的那万亿次无声而同步的闪烁，是如何让一个人得以思考、感知、记忆、想象、创造以及感觉。大约与此同时，我还在阅读研究大脑的著作，研习数门外语，琢磨来自不同国家的异域书写系统，想尽办法让一台计算机生成语法复杂而近乎连贯通顺的英语或其他语言的句子，并选修了一门格外吸引人的心理学课程。所有这些不同的路径，都汇聚在一片浓稠的问题星云之上，盘绕在思维与机械结构以及思维状态（mentality）与机械性之间。

于是，有关模式的研究（数学）与有关悖论的研究（元数学）便在我青春期的头脑里缠结在了一起。我莫名地产生了一种信念，认定但凡我能深刻地把握住这两门彼此纠缠的学科，所有那些令我执迷的神奇秘密便可以一清二楚地展现在我的眼前。虽然在随后的20年间，我渐渐地抛弃了自己的信仰，直到完全不再相信这些学科里包含（哪怕是隐含）了所有问题的答案；但还有一样东西我从未曾丢弃，那便是我那个灵感式的直觉围绕着“我是什么？”这一永恒谜题的核心，哥德尔精心打造的那个缠结的结构，正轻盈缥缈地涡旋。

正因如此，虽然在这本书中主要驱动我前进的，是关于意识和自我的问题，可我还是不得不用几页文字，（粗略）交代理解哥德尔的观念——这里特别指数字理论与逻辑——所必需的背景。当然，对于这两个领域的介绍，都不需要浓墨重彩，但我至少也要粗线条地勾勒出它们基本上是怎么回事；不然的话，我们就没办法继续下去了。所以，亲爱的读者，请系紧你的安全带。我们在接下来的两章里，要遭遇气流，经历一点儿颠簸了。

附言

在心满意足地写完了这一章之后，我忽然想起自己还买过两本关于“有趣数字”的书——《古怪而有趣数字的企鹅词典》（*The Penguin Dictionary of Curiousand Interesting Numbers*），作者戴维·韦尔斯（David Wells）是一位我十分钦佩的数学家，《非凡的数字》（*Les nombres remarquables*），作者弗朗索瓦·勒利奥内（François Le Lionnais）是法国著名的文学运动“乌力波”两位发起人之一。我还能隐隐约约地记得，这两本书都按照数值大小的顺序列出了它们的“有趣数字”，所以我决定把它们找出来看看，被它们各自排除在外的最小整数是哪个。

正如我所料想的一样，两位作者都做出了相当感人的努力，想要囊括所有存在的整数，但人类的知识毕竟是有限的，而人类的生命也有尽头，所以不可避免地，每本书都或早或晚地开始跳过了一些数字。韦尔斯跳过的第一个数字是43，勒利奥内坚持得稍微久了一点，推迟到了49。我个人对于43被跳过并不感到特别惊讶，但是对于49被跳过颇感意外；它毕竟是一个平方数，多少也该有一点有趣性。可另一方面，我也愿意承认，当你遇到过好几个平方数之后，平方数这个特征会变得无聊起来，所以我可以部分地理解为什么49单凭自己平方数的特性不足以跻身勒利奥内的最终名单。韦尔斯为49举出了几条有趣的属性（但是没有包括它是平方数这一点），而反过头来，勒利奥内又指出了43所具备的几点十分出人意料的性质。

所以接下来，我决定找到那个被两本书同时认定为完全无趣的最小数字，结果是62。容我冒昧一提，这本书的英文版首次出版的时候，我刚刚满62岁。有没有可能，62说到底也是一个有趣的数字呢？

---

1. 有很长一段时间，我都深爱着艾舍尔的艺术，但随着时光流逝，我发现自己越来越被他早期那些非悖论性的风景画所吸引了，我在那些平凡的场景中感受到了他无处不在

的魔力。可参见我在2002年写过的一篇纪念艾舍尔诞辰100周年的文章。

2. Three excellent books on paradoxes are [Falletta], [Hughes and Brecht], and [Casati and Varzi 2006].

3. 在《数学原理》中，只有两个人得到了（几乎）自给自足的作者们的感谢，贝里就是其中的一位。

4. 正多面体只有5种：正四面体、正六面体、正八面体、正十二面体和正二十面体。——编者注

5. 这类数字又称为完全数，6是第一个完全数。——编者注

6. 儿童黑话（Pig Latin）是一种英语隐语，形式是在英语上加上一点规则使发音改变。——编者注

7. See [Chaitin], packed with stunning, strange results.

第9章
模式与可证性

《数学原理》及其定理

在20世纪早期，罗素秉持过这样一个信条："发现悖论，研究悖论；设计并筑起把它们挡在墙外的伟大壁垒！"（不过这句话是我重述的，不是他的原话。）《数学原理》便是他为数学推理新建的一座屏障重重的堡垒。在这个信条的鼓动下，他断然认定，任何集合都不可包含其自身，而任何句子也都不可反身谈论它自己。这个双重禁令的并行，意在保护《数学原理》免于落入那些更加天真的理论曾堕入的陷阱。可是，当库尔特·哥德尔对于被我称为*PM*——就是在《数学原理》中用于推理集合（后来也用于有关数的推理，只不过是以集合形式定义的数字）的形式系统（formal system of *Principia Mathematica*）——展开仔细的审视时，却发生了某些特别奇怪的事情。

请允许我先对《数学原理》和*PM*之间做一点更加明白的区分。前者是一套三卷本的鸿篇巨制，而*PM*则是在这套书中以艰深晦涩的符号（见本章末）呈列出来并深入探讨的一套精准的符号操作法则。这种区分可以类比于伊萨克·牛顿（Isaac Newton）那本名为《自然哲学的数学原理》的巨著与他在那本书中阐述的力学定律之间的区别。

虽然说，即便如"1加1等于2"（用*PM*的符号表示为[注]"s0+s0=ss0"，其中字母"s"代表的是"后继"的概念）这样一个相当基本的事实，要想使用*PM*调度符号的严格法则得出严谨的论证，也需要不下数个章节的定理与推导，可是哥德尔仍然意识到了如此笨拙繁重的*PM*在言说自然数

时所具有的强大力量——事实上，其言说的是自然数任意微妙的性质。（顺便一提，在“任意微妙的性质”这个小小的短语中已经暗藏了天机，只不过在这天机之上，蒙着一层厚厚的面纱，以至于几乎无人能够察觉到，这几个字里究竟有多少含义。而哥德尔一眼洞穿了它。）

例如，只要我们在《数学原理》中引入足以容纳加减法等基本算术运算的集合理论装置，那么在*PM*形式主义的内部就可以很容易定义出某些更加有趣的概念了，比如“平方数”（即一个自然数的平方）、“非平方数”、“质数”和“合数”。

因而至少从理论上来说，可以用一卷《数学原理》来专门探讨哪些整数是两个平方数之和[注]而哪些整数不是的问题。例如，41就是16和25的和，而且还有其他数不清的整数，是两个平方数相加得到的结果。我们把它们归为A类的成员。另一方面，43则不是任何一对平方数之和，而且同样也有不计其数的整数，不能由两个平方数相加得到。我们把这些数字称为B类的成员。（109是哪一类的成员？133呢？）深入而透彻地理解这种全部整数集合的优雅二分法虽然是一项极其微妙难解的任务，可是数论学家们早在哥德尔出生以前就已经大功告成了。

类似地，我们可以想象另一卷《数学原理》，其专门探讨的问题是：哪些整数是两个质数之和[注]，而哪些整数不是。例如，24是5和19相加之和，而23则不是任何两个质数相加之和。这一次，我们可以把这两类整数分别归为“C类”和“D类”。每一类整数都有无穷多的成员。深入而透彻地理解这一种全部整数集合的优雅二分法，对数论学家们来说是一项严峻的挑战；虽然这个问题自提出以来，已经在两个多世纪里获得了不少的推进，可至今仍是一个未解的难题。

混合两种不太相干的观念：质数与平方数

在我们正式投入到哥德尔有关*PM*的那个出人意料而基于扭转的洞见之前，我需要首先讲一讲发现模式的美妙乐趣，然后再讲一讲理解隐含在模式背后之原理的美妙乐趣。正是数学家们对于为什么的不懈探索，定义了他们所属学科的本性。我希望可以借助我在数论里最喜欢的一个事实，以一种愉快欣悦的方式对此加以阐明。

让我们先问自己一个有关质数的足够简单的问题：哪些质数是两个平方数之和（例如，41）而哪些质数不是（例如，43）？换句话说，让我们返回A类与B类这两个无限的集合之中，问一问它们各自包含哪些质数。有没有可能，绝大多数的质数都属于其中一类，而只有少数例外在另一类之中？或者，分布比例是一半对一半？在每一类中都有无限多个质数吗？给出一个任意的质数*p*，有没有一种简单快捷的方式，可以检验出*p*属于哪个类别（而不是试遍所有比*p*小的两个平方数的和）？质数在这两种类型中的分布是遵循着一种可预测的模式，还是毫无头绪，一团乱麻？

对于某些读者而言，这些问题似乎太过古怪，甚至违反自然，难以处理，可是，数学家们天生就是一种充满好奇的动物，他们又恰好经常深深地着迷于探索两个似乎先在没有任何关系的两个概念（如质数和平方数）之间的交互关系。通常的结果是，某种出乎意料而又相当紧密的关系浮出了水面——那种离奇的隐秘规律感觉就像魔法一样，发现它，或者说揭示它，甚至能引起一个人沿着脊柱上下传递的阵阵战栗。而我作为其中的一员，也要厚着脸皮承认，自己对于这种令人脊柱发麻的敬畏、美妙、神秘与惊喜的混合物，具有一种高度的敏感性。

为了找到一点这种感觉，让我们先来列举出小于100的所有质数——2，3，5，7，11，13，17，19，23，29，31，37，41，43，47，53，59，61，67，71，73，79，83，89，97。不得不说，这是一个相当混乱无序的清单。然后，我们再重新呈现一次这份清单，把那些是两个平方数之和的质数（也就是属于A类的质数）加粗，而让那些不是两个平方

数之和的质数（也就是属于B类的质数）保持不变。我们将会得出如下的清单：

2，3，**5**，7，11，**13**，**17**，19，23，**29**，31，**37**，**41**，43，47，**53**，59，**61**，67，71，**73**，79，83，**89**，**97**，…

你能在这里看出有任何有趣的事情发生吗？别的不说，两类数字的分布基本上不相伯仲，这是不是已然算是一个不小的惊喜了呢？为什么会是这样一种情况呢？为什么不论是A类还是B类都没能独大？在继续向后推进一段之后，A类质数或B类质数会不会逐渐占据主导，还是大体上永远保持势均力敌的态势？当我们朝着无限的方向不断推进时，这种平衡的态势会越来越趋近于精确的对半开吗？如果当真如此的话，为什么会出现这样一种令人惊叹的精致的平衡呢？对于我而言，这其中就存在了某种具有巨大诱惑力的东西，所以我也建议你在继续往下阅读之前，多端详一会儿——几分钟就好——这个数列，努力找找，看其中是否有任何模式可循。

探索模式

很好，读者，我希望你已经独立探索了上面的模式，现在让我们重新回来。你很可能已经注意到，我们通过加粗这一动作，似乎在无意之中、纯属偶然（或者这是偶然吗？）地把这份清单分割成了诸多的单个数和成对数。这是不是揭示出了某种隐含的联系呢？

让我们循此思路更进一步。加粗的成对数是13-17，37-41和89-97，而未加粗的成对数是7-11，19-23，43-47，67-71和79-83。接下来，我们设定用字母“P”代替每一组成对数，而用字母“S”表示每一例单个数，同时保留区分A类与B类的加粗标记。如此一来，我们就能得到如下的

字母序列：

S，S，**S**，P，**P**，P，**S**，S，**P**，P，**S**，S，**S**，P，**S**，P，**P**，…

这里面存在某种模式吗？还是根本没有模式可言？你怎么想？如果我们抽取出其中属于A类的字母，就会得到：**SSPSPSSSP**；而如果我们只抽取出属于B类的字母，则会得到：SPPSPSPP。如果这里存在某种周期性规律或者更加微妙的节奏类型，那么也一定不是可以轻易发现的。不管是在加粗的序列、非加粗的序列还是它们混合在一起的序列中，都没有任何简明的可预测的模式一目了然地跃然纸上。我们只是隐约地感受到，在两种类型之间，有一种相当平衡的均势，但是我们完全不清楚其原因何在。这既惹人心痒，又催人憔悴。

不屈不挠地探索模式之人

到了这个节骨眼上，我觉得自己最有必要指出的已经不是两类数字之间的区别，而是两类人之间的区分了。有些人立刻就会被探索模式这一想法所吸引，而有些人则会觉得它没什么意思，甚至倒人胃口。从本质上说，前者就是有数学倾向的人，而后者则不是。发现那些看起来原本并不存在的模式，正是身为数学家的那些人在内心深处难以抗拒——事实上很容易就被其点燃——的冲动。正是在表面的无序中对于有序的热切追寻，点燃了他们的火光[注]，并燃烧着他们的灵魂。亲爱的读者，我希望你也在这样一类人中间，但即便你不在此类，也请容我把下面这一段讲完。

我们似乎已经钦定了一种归类模式，即我们将永远只会遇到单个数和成对数。即便我们无法确切地说出P和S们散布的方式，但至少表面看起来，“两个平方数之和对非两个平方数之和”的奇妙二分法加诸质数序列之上，又分化出单个数与成对数两种类型，这已经是一个相当神奇的

发现了！谁猜得到这一点呢？

遗憾的是，我现在必须坦白交代，其实是我误导了你。我们只需要把下一个质数放到我们的清单里，也就是101，我们发现的那个似乎成立的秩序就已经被打破了。别忘了，质数101可是两个平方数1和100的和，属于A类，也必须标为加粗，所以我们先前所称的加粗成对数89-97其实是一个加粗的三连数。如此一来，那个单纯由S和P构成的序列，那个满载着我们希望的构想，便成了竹篮打水一场空。

一位模式的探索者在这时会采取什么行动——放弃吗？当然不是！在经历了一次挫败之后，韧性十足的模式探索者想必一定会重整旗鼓。没错，就像刚刚这个成语给予我们提示的一样，让我们换不同的方式来重新梳理一下质数的序列。让我们把两类数字分梳区隔，呈列独立的两行。我们将得到如下的数列：

是 平方数加平方数：2，5，13，17，29，37，41，53，61，73，89，97，101…

非 平方数加平方数：3，7，11，19，23，31，43，47，59，67，71，79，83…

你看出什么来没有？如果没有，让我给你一点提示。如果简单地把每一行中的相邻两个数字求差，会得到什么？你自己来试一试——或者，如果你很懒的话，就继续往后读吧。

在上面那一行中，你会得到3，8，4，12，8，4，12，8，12，16，8，4；而在下面一行中，你得到的是4，4，8，4，8，12，4，12，8，4，8，4。进行到这一步，哪怕是再漠不关心的读者也一定有所发现了：少数几个整数（4，8和12）比重最大，不仅如此，这几个整数还全部是4的倍数。这看起来已经绝不能用单纯的巧合来解释了。

而在这两份清单中唯一的一个最大的数字，16，也是4的倍数。这种新的模式——即仅包含4的倍数——能够永远保持下去吗？（当然，第一行最开始有一个扫兴的3，但是我们可以把这归结于2是唯一的偶数质数这个事实之上。没什么大问题。）

有模式，即有理由

上文那几行话里包含的核心思想是这种模式不可能仅仅是一种巧合。一位发现了此类模式的数学家会本能地发问："为什么？这种秩序背后的理由是什么？"非但所有的数学家都想知道那个理由是什么，而且更为重要的是，他们都内在地相信，不管有没有人找到了理由，那个理由一定存在。在数学的世界里，没有任何事是"碰巧"发生的(注)。一种完美模式、一种永远成立的规律性的存在，就好像浓烟揭示了火焰一样，揭示出幕后必然有某事正在发生。寻找这样东西，揭开它的面纱，将它公之于众，被数学家认为是自己神圣的目标。

如你所知，这就是被称为"寻找证明"的行动，或者换个说法，就是把一个猜想转变为一个定理的行为。最近故去的一位伟大而乖僻的匈牙利数学家保尔·厄尔多斯（Paul Erdös）曾滑稽地评论道(注)："数学家就是把咖啡转换成定理的机器。"虽然他的这句俏皮话里自然不乏真相，但是更准确地说，数学家应该是发现猜想并把它们转换为定理的机器。

构成数学思维基础的是一个雷打不动的信念，即一旦某个数学陈述X为真，则X必有一个证明，反之亦然。确实如此，在数学思维中，"拥有证明"和"为真"具有完全等同的意义！反过来说，"为假"也就意味着"没有证明"。人们可以像我们在前面做过的一样，通过数字上的探索，发现关于一个完美而无限的模式的提示；可是，人们怎么才能确定地知道，一种猜想中的规律性可以无穷无尽地持续下去呢？例如，人们如何才能确定，质数的个数是无限的呢？我们怎么能确定，永远不会

在某一点遇到最后一个质数——最大质数*P*呢？

如果*P*真的存在，那么它一定是一个非常重要而且相当有趣的数字；可是，只要你看过一份足够长的连续质数清单（前文那个100以内的清单就能给出大概的意思了），就会发现，虽然它们的节奏有一点“颠簸不平”，零星散布着不规则的间隔，但这些质数之间的间隔相比于其两侧的质数而言，总是相当之小。照此十分清晰的趋势来看，如果突然之间质数穷尽了，那么这感觉就好像在没有任何警示的情况下从地球的边缘跌落了一样。那将是一次极大的冲击。可是话又说回来，我们怎么知道这不会发生呢？或者说，我们真的知道吗？在一台计算机的帮助下，我们可以发现，质数在10亿、万亿的级别上仍在不断涌现，这的确很棒，可是也不能板上钉钉地保证它们不会在更远的某处陡然停止。要想得到确切的答案，我们不得不依赖推理，因为，虽然有限数量的证据可以给我们很强的提示，但它们就是不合格，因为在无限与任何一个有限的数字之间，都存在本质上的不同。

航行于质数的海洋，并于边缘坠落

你可能已经在什么地方看到过欧几里得（Euclid）对于质数无限性的证明，但如果你没有的话，你错过的便是人类知识有史以来发现的最为关键的理论之一。这将成为你生命经验中的一个空白，就像从未品尝过巧克力或者从未听过一曲音乐那样令人遗憾。我无法容忍自己的读者在知识上存在如此致命的空白，所以也管不了那么多了！

让我们假设那个普天之下最大的质数*P*真的存在，然后看看这个预设会将我们引向何方。若*P*存在，则意味着有一个包括全部质数的有限而封闭的俱乐部，而*P*自己则是那位荣耀的、冠冕加身的终极成员。下面，让我们斗胆把这个封闭俱乐部里所有的质数连乘起来，得到一个庞大到离谱的数字*Q*。于是，这个数字*Q*便可以被2整除，也能被3，5，

7，11等等质数整除。根据它的定义，Q可以被俱乐部里的每一个成员——即全宇宙中的每一个质数——整除！下面，让我们完成画龙点睛的一笔，就像在生日宴会上点亮最后一根蜡烛那样，让Q再加上1，得到$Q+1$。我们在此得到了一个确定其<u>不是</u>质数的庞大数字，因为P（显然比Q小得多）才是伟大的最后的质数，是所有质数里最大的一个。根据我们最初的设定，所有比P更大的数字都是合数。因此，远比P大的$Q + 1$便也应该是一个合数，也<u>不得不</u>具有某个质数因子。（请务必将此点铭记于心。）

那个未知的质数因子能是什么呢？它不可能是2，因为2可以整除Q，Q只比$Q+1$小1，而两个偶数之间的差距绝不可能是1。它也不可能是3，因为3也可以整除Q，而能被3整除的数字也绝不可能是相邻的！事实上，不管我们从哪个俱乐部里面选择了哪个质数P，我们都会发现，P不可能整除$Q+1$，因为P可以整除与它相邻的较小数字Q（而P的倍数绝不可能彼此相邻——它们每隔P个数字才出现一次）。所以，推理告诉我们，<u>没有任何一个</u>属于有限而封闭的质数俱乐部的成员可以整除$Q+1$。

但我在更早的时候已经得出（而且我也请你铭记过），$Q+1$既然是一个合数，就<u>不得不</u>具有一个质数因子。叮！我们掉落陷阱，把自己逼入了绝境。我们生造出一个疯狂的数字——这个数字一方面必须是一个合数（即具有某些比自身更小的质数因子），另一方面则没有任何比自身更小的质数因子。这种自相矛盾发生的来源，在于我们预设了一个以P为其冠冕的有限而封闭的质数俱乐部，所以我们除了回头擦除一整套悦目而可疑的幻景以外，别无选择。

不可能存在一个“普天之下最大的质数”；也不可能存在一个“包含全部质数的有限而封闭的俱乐部”。那些都是虚构的。而真相则如我们刚刚阐明的那样，质数的清单将无限延长，无穷无尽。不管我们走多远，也永远、绝对不会“跌落出地球”。我们如今的确信来自天衣无缝的

推理，这是任何航行于数字海洋之中的有限的计算都无法给予的确信。

如果这有幸成为你理解为什么不存在最大的质数（而不是仅仅知道是这么回事儿）的初次经验，我希望你能够像品尝一块巧克力或品味一段音乐一样对它好好加以体会。同那些经验一样，玩味这番证明也是一种快感的来源，你可以重复多次地品味并沉浸其中，每一次都会感觉耳目一新。另外，这份证明还是引发其他证明的丰富源泉——欧几里得的主题变奏(注)（但我们就不在此展开探讨了）。

数学家的信条

我们刚才的详细所述，正是我所谓“数学家的信条”的一则可爱范例。我把这个信条总结如下：

X为真，因为X有证明；

X为真，所以X有证明。

请注意，这是一条双行线。信条的前半部分声明，证明是真理的保证，而后半部分则声明，有规律，即有理由。当然，我们自己也许发现不了隐藏的理由，但是我们坚定不移地相信理由是存在的，而且原则上总可以在某一天由某一个人发现。

对于一位数学家而言，质疑信条的任何一部分都是不可思议的。质疑第一行信条，就要想象出一个已被证明的陈述依然为假的情况，这是对“证明”这一概念活生生的嘲讽；而质疑第二行则要想象出一种在数学领域内完美的、绝无例外的模式，它永远都成立，却没有任何节奏或理由。在数学家们看来，这种既无瑕疵又无理由的结构根本就是一种说不通的观念。在这一点上，所有的数学家都算得上爱因斯坦的同道，因为后者的一则著名言论正是：“上帝不掷色子。”(注)爱因斯坦的意思是，从

本质上说，没有任何一件事情的发生是没有原因的；对于数学家们而言，一致的、潜在的原因永远存在，这是他们不可动摇的信念。

没有什么无限的巧合

我们现在再说回A类与B类质数，因为我们还没有完全抵达真相，尚未经历我说过的那种神秘的战栗之感。还记得吗？我们已经注意到，两行的特征都是，每两个数字之间的差可以用$4n$的形式表示——也即是说，4，8，12，如此类推。我们没有证明这个事实，但是我们通过足够多的观察，做出了这一猜想。

在我们所列的下面一行中，第一个数字是3，所以我们的猜想就是，那一行中所有的数字都可以通过在3之上加上4的倍数来得到，结果就是，那一行中的每个数字都可以写成$4n+3$的形式。同理（如果我们忽视最开始不符合规律的2），上面一行中的第一个数字是5，所以如果我们的猜想是正确的，那么那一行后面的每一个数字都可以写成$4n+1$的形式。

不错——我们的猜想为我们提出了一个格外简单的模式：满足$4n+1$形式的质数可以呈现为两个平方数之和，而$4n+3$形式的质数则不能。如果这种猜测没错的话，它就在质数与平方数（两种似乎没有任何先在联系的数字类型）之间建立起了一种美妙而惊人的联系，打我们一个措手不及。这是投向纯正魔法的一瞥——那种数学家们赖以为生的魔法。

然而对于一位数学家而言，这种一闪而过的欢欣，只不过是故事的开端而已。这就像一场谋杀悬疑案：我们已经发现某人身亡，可是凶手是谁呢？总归要有一个解释才行。找到或者理解真相或许并没有那么简单，但真相必然存在。

在此，我们知道（或者至少强烈地怀疑）有一种美妙的无限的模式

存在，但是它存在的理由是什么？最基本的假设是，这里有一个理由——我们的模式远非一种“无限的巧合”，而是源自某个单一的、强有力的潜在理由；在所有这些大量的“独立”事实的背后，其实只存在着一种现象。

事实上，我们瞥见的模式恰巧还有更加丰富的意味。不仅满足$4n+3$形式的质数绝不可能是两个平方数之和（证明这一点是很简单的），而且原来每个满足$4n+1$形式的质数有且只有一种写成两个平方数之和的形式。以101为例。不仅101等于100 + 1，而且没有任何另外两个平方数相加的和还等于101。最后，当推至极限时，A类数与B类数的比例逐渐趋近于1。这就意味着，我们在100以内的质数中观察到的精致平衡和我们对这种平衡将永远持续下去的猜想，得到了严格的证明。

虽然我不会更深入地探讨这个特别的案例，但是我要指出的是，很多数论教科书都证明了这个（绝非微不足道的）定理[注]，由此为一种模式补充了一个证明。正如我在先前所说，X为真，是因为X具有一个证明，而反之亦然，X为真，因此X具有一个证明。

寻找证明及其本性的漫漫求索

“哪些数字是两个质数的和？”我在前面提到的这个问题，首次被人提出是在近300年之前，而至今也没有得到圆满的解决。然而，数学家们都是矢志不渝的求索者，而他们对于一个证明的求索可以持续几个世纪，甚至上千年。他们在为一个从数字趋向上来看似乎永远成立的数学模式寻找证明的时候，即便经历了成千上万次的失败，也决不却步。的确，对于绝大多数人而言，一种数学猜想在实际经验中得到广泛的确认，这就已经可以令人满足了，可是对于数学家而言，这只会让他们更加兴奋而抓狂。他们想要的是一个证明，与欧几里得的证明一样坚实可靠，而不是大量的抽样调查而已！而驱动他们的信念是：这个证明必然

存在——换言之，若证明不存在，则所言模式一定为假。

于是，这便构成了数学家信条的反面：

X为假，因为X没有证明；

X为假，所以X没有证明。

一言以蔽之，正如可证性与真理在数学家眼中同为一物，不可证性与虚假也是如此。它们是同义词。

在文艺复兴之后的几个世纪里，数学分化出很多不同的分支学科，而不同的分支中也出现了各种各样不同的证明。然而，间或会有某些明显荒谬的结论似乎得到了严格的证明，而又无人能够指出哪里出现了差池。而当越来越奇怪的结论轮番现身之后，证明本身的不确定性变得越来越令人不安，直到19世纪中叶，一场强有力的运动终于爆发了，其目标正是要敲定推理究竟为何物，并将其永远与数学绑定在一起，合二为一。

很多哲学家和数学家都为这一高尚的目标做出了贡献，而在上个世纪之交，这个目标似乎已经近在咫尺。数学推理似乎已经被精确地定义为对于某些基本逻辑法则的重复运用，这些逻辑法则又名推理规则，如肯定前件式（modus ponens）：若已证明结论*X*，并同时证明*X*=>*Y*（箭头表示蕴涵关系，所以这个表达式意味着“若*X*为真，则*Y*亦为真”），那么*Y*即被证明。除此之外，还有一些其他的基本推理规则，但是获得公认的是，这样的规则不需要太多。大概到了20世纪的第一个10年[注]，罗素和怀特海把这些规则编成了统一且相当麻烦的符号（见下一页），由此明确地把所有不同的数学分支都纳入了逻辑之内，形成了一个密不透风的完美统一体。

罗素和怀特海的大作《数学原理》让人们不再需要担心落入虚假推

理的阴沟。定理被简单地理解为符号操作序列的底线，而其顶线则是公理或更早的定理。数学真理如此优雅利落地聚合在一起。而正当这只圣杯仿佛触手可及之际，一个年轻的男孩子也在奥匈帝国的布尔诺（Brünn）[注]成长了起来[注]。

$*97{\cdot}31.\quad \vdash . (\overrightarrow{B}\text{‘}R) \uparrow \mathrm{Cnv}\text{‘}\overleftarrow{R_*} \in \epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R . \mathrm{D}\text{‘}\{(\overrightarrow{B}\text{‘}R) \uparrow \mathrm{Cnv}\text{‘}\overleftarrow{R_*}\} = \overrightarrow{B}\text{‘}R$

Dem.

$\vdash . *97{\cdot}3 . *85{\cdot}13 \dfrac{\breve{R}_*}{Q} . \supset$

$\vdash : S \in (\breve{R}_*)_\Delta\text{‘}\overrightarrow{B}\text{‘}R . \supset . S \,|\, \mathrm{Cnv}\text{‘}\overleftarrow{R_*} \in \epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R \qquad (1)$

$\vdash . (1) . *97{\cdot}301 . \quad \supset \vdash . I \upharpoonright \overrightarrow{B}\text{‘}R \,|\, \mathrm{Cnv}\text{‘}\overleftarrow{R_*} \in \epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R .$

$[*50{\cdot}61] \qquad \supset \vdash . (\overrightarrow{B}\text{‘}R) \uparrow \mathrm{Cnv}\text{‘}\overleftarrow{R_*} \in \epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R \qquad (2)$

$\vdash . *35{\cdot}62 . *33{\cdot}431 . \supset \vdash . \mathrm{D}\text{‘}\{(\overrightarrow{B}\text{‘}R) \uparrow \mathrm{Cnv}\text{‘}\overleftarrow{R_*}\} = \overrightarrow{B}\text{‘}R \qquad (3)$

$\vdash . (2) . (3) . \supset \vdash . \mathrm{Prop}$

$*97{\cdot}32.\quad \vdash . \overrightarrow{B}\text{‘}R \in \mathrm{D}\text{“}\epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R \quad [*97{\cdot}31]$

$*97{\cdot}33.\quad \vdash : R \in 1 \to 1 . \alpha \subset s\text{‘}\overleftrightarrow{R_*}\text{“}\beta . \beta \subset s\text{‘}\overleftrightarrow{R_*}\text{“}\alpha . \supset . \overleftrightarrow{R_*}\text{“}\alpha = \overleftrightarrow{R_*}\text{“}\beta$

Dem.

$\vdash . *97{\cdot}15 . \mathrm{Fact} . \supset \vdash :. \mathrm{Hp} . \supset : y \in \beta . x \in \overleftrightarrow{R_*}\text{‘}y . \supset . \overleftrightarrow{R_*}\text{‘}y = \overleftrightarrow{R_*}\text{‘}x . y \in \beta .$

$[*37{\cdot}62] \qquad \supset . \overleftrightarrow{R_*}\text{‘}x \in \overleftrightarrow{R_*}\text{“}\beta \qquad (1)$

$\vdash . (1) . *10{\cdot}11{\cdot}21{\cdot}23 . *40{\cdot}4 . \supset \vdash :. \mathrm{Hp} . \supset : x \in s\text{‘}\overleftrightarrow{R_*}\text{“}\beta . \supset_x . \overleftrightarrow{R_*}\text{‘}x \in \overleftrightarrow{R_*}\text{“}\beta :$

$[\mathrm{Hp.Syll}] \qquad \supset : x \in \alpha . \supset_x . \overleftrightarrow{R_*}\text{‘}x \in \overleftrightarrow{R_*}\text{“}\beta :$

$[*37{\cdot}61] \qquad \supset : \overleftrightarrow{R_*}\text{“}\alpha \subset \overleftrightarrow{R_*}\text{“}\beta \qquad (2)$

$\vdash . *40{\cdot}4 . \supset \vdash :. \mathrm{Hp} . \supset : y \in \beta . \supset . (\exists x) . x \in \alpha . y \in \overleftrightarrow{R_*}\text{‘}x .$

$[*97{\cdot}15] \qquad \supset . (\exists x) . x \in \alpha . \overleftrightarrow{R_*}\text{‘}x = \overleftrightarrow{R_*}\text{‘}y .$

$[*37{\cdot}62] \qquad \supset . \overleftrightarrow{R_*}\text{‘}y \in \overleftrightarrow{R_*}\text{“}\alpha \qquad (3)$

$\vdash . (3) . *37{\cdot}61 . \supset \vdash : \mathrm{Hp} . \supset . \overleftrightarrow{R_*}\text{“}\beta \subset \overleftrightarrow{R_*}\text{“}\alpha \qquad (4)$

$\vdash . (2) . (4) . \supset \vdash . \mathrm{Prop}$

$*97{\cdot}34.\quad \vdash : R \in 1 \to 1 . \beta \in \mathrm{D}\text{“}\epsilon_\Delta\text{‘}\overleftrightarrow{R_*}\text{“}\alpha . \supset . \overleftrightarrow{R_*}\text{“}\alpha = \overleftrightarrow{R_*}\text{“}\beta$

Dem.

$\vdash . *83{\cdot}6{\cdot}62 . \qquad \supset \vdash :. \mathrm{Hp} . \supset : x \in \alpha . \supset_x . \exists ! \beta \cap \overleftrightarrow{R_*}\text{‘}x : \beta \subset s\text{‘}\overleftrightarrow{R_*}\text{“}\alpha \qquad (1)$

$\vdash . *40{\cdot}4 . *97{\cdot}101 . \supset \vdash :. x \in \alpha . \supset_x . \exists ! \beta \cap \overleftrightarrow{R_*}\text{‘}x . \equiv . \alpha \subset s\text{‘}\overleftrightarrow{R_*}\text{“}\beta \qquad (2)$

$\vdash . (1) . (2) . *97{\cdot}33 . \supset \vdash . \mathrm{Prop}$

$*97{\cdot}341.\quad \vdash : R \in 1 \to 1 . \beta \in \mathrm{D}\text{“}\epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R . \supset . \overleftrightarrow{R_*}\text{“}\beta = \overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R$

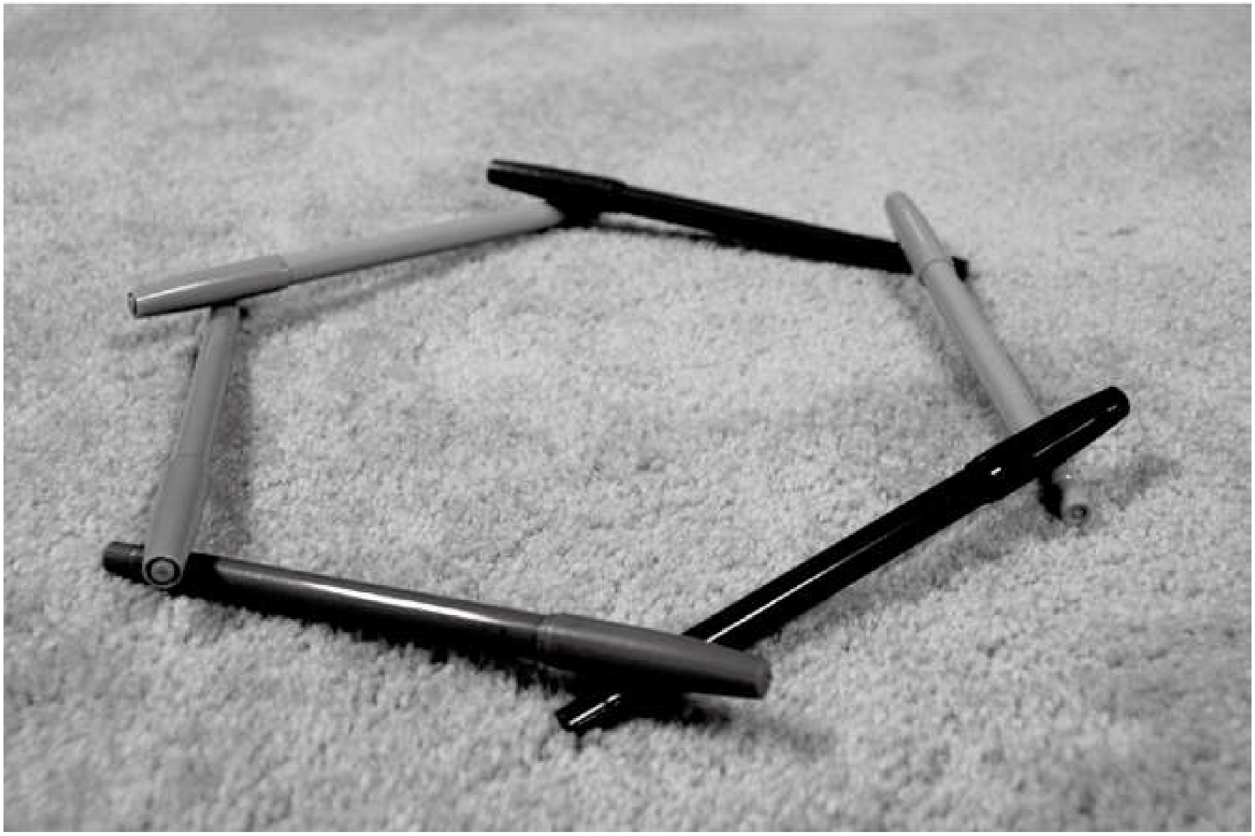
$\left[*97{\cdot}34 \dfrac{\overrightarrow{B}\text{‘}R}{\alpha} . *97{\cdot}2\right]$

$*97{\cdot}35.\quad \vdash : R \in \mathrm{Cls} \to 1 . T \in \mathrm{Potid}\text{‘}R . \overrightarrow{B}\text{‘}R \subset \mathrm{D}\text{‘}T . \supset .$

$\mathrm{Cnv}\text{‘}\{(\overleftarrow{R_*} \upharpoonright \overrightarrow{B}\text{‘}R) \,|\, T\} \in \epsilon_\Delta\text{‘}\overleftarrow{R_*}\text{“}\overrightarrow{B}\text{‘}R . \mathrm{G}\text{‘}\{(\overleftarrow{R_*} \upharpoonright \overrightarrow{B}\text{‘}R) \,|\, T\} = \breve{T}\text{“}\overrightarrow{B}\text{‘}R$

Dem.

$\vdash . *97{\cdot}3 . *92{\cdot}101 . \supset \vdash : \mathrm{Hp} . \supset . \mathrm{Cnv}\text{‘}\{(\overleftarrow{R_*} \upharpoonright \overrightarrow{B}\text{‘}R) \,|\, T\} \in 1 \to \mathrm{Cls} \qquad (1)$



1. 我在此借用了哥德尔对于*PM*记号系统的简化版本，而没有直接使用原来的系统，因为后者是在是太难消化了。（翻到本书第147页看一下，你就明白我的意思了。）
2. See [Hardy and Wright] and [Niven and Zuckerman].
3. See [Wells 2005], an exquisite garden of delights.
4. See [Ulam], [Ash and Gross], [Wells 2005], [Gardner], [Bewersdorfi], and [Livio].
5. See [Davies].
6. Erdös, a devout matheist, often spoke of proofs from “The Book”, an imagined tome containing God’s perfect proofs of all great truths. For my own vision of “matheism”, see Chapter 1 of [Hofstadter and FARG]
7. See [Chaitin].
8. See [Hoffmann], one of the best books I have ever read
9. See [Hardy and Wright] and [Niven and Zuckerman].
10. The history of the push to formalize mathematics and logic is well recounted in [DeLong], [Kneebone], and [Wilder].

11. 布尔诺现在属于捷克。——编者注

12. See [Goldstein] and [Yourgrau].

第10章
哥德尔的典型怪圈

当哥德尔遭遇斐波那契数列

在20岁刚出头的年纪，那位来自布尔诺的男孩子已经成为一位超凡的数学家了，而且和所有的数学家一样，他也知道自然数拥有无穷无尽的变体。除了平方数、立方数、质数、10的幂、两个平方数之和以及所有其他常见的候选者以外，他还熟稔很多整数的类型。而对年轻的库尔特的未来至关重要的是，他在列奥纳多·迪比萨（Leonardo di Pisa，更为人熟知的名字是“斐波那契”）的启发下，明白了可以用递归的方式来定义整数的种类。

斐波那契在14世纪构造并探讨了我们如今所说的“斐波那契数”[注]：

1，2，3，5，8，13，21，34，55，89，114，233，377，610，987，1 597，…

在这一急速增长的无穷数列中，每一个新的元素都是由前面两个数相加得到的，而我在此后提到这些数字时，会称之为*F*数（只有前两个数，1和2，是不符合规律的，我们强行规定它们为*F*数）。

这种以其自身定义数列的近乎循环而并非完全循环的方式[注]，被称为“递归式定义”。这意味着存在某种可以从前面数字推导出后面数字的精确计算规则。这种规则可能包括加、减、乘、除，随便什么都行——只要定义明确即可。可以把一个递归序列的开局（在这里，就是数字1

和2）想象成一包种子，由此发端，基于固定的规则、按照预定的方式，将生长成一株巨大的植物——枝繁叶茂，无穷无尽。

里海的宝石：一则寓言

列奥纳多·迪比萨的数列充溢着神奇的模式，遗憾的是，投身其中只会让我们越来越偏离航线。尽管如此，我还是忍不住要指出，在这份清单上的前几个*F*数字中，144脱颖而出，因为它是一个显眼的完全平方数。除了8是个立方数，还有1是个相当特殊的退化情况之外，在*F*数列的前100项里，就没有任何完全平方数、立方数或者任何数次幂了。

几十年以前，人们开始追问，8和144在*F*数列中的出现，究竟是出于某种理由，还是只不过是一个“随机的意外”。因此，随着计算工具日益强大起来，探索的任务就落在了它们的肩上。奇妙的是，即便在超级计算机出现之后，数以百万乃至10亿计的*F*数可以被快速地计算出来，可是仍然没有任何人在斐波那契数列中遇见过任何其他的完全幂。在*F*序列中很快找到一个整数幂的希望看起来越来越渺茫了，但是为什么会发生这样一种完全的相互排斥呢？某个任意的整数n的n次幂，与斐波那契数列以奇异的递归形式把相邻的两个数相加，存在什么关联呢？8和144会不会仅仅是两个随机的小故障呢？为什么没有其他的小故障发生呢？

我们可以用一则寓言来阐明这个道理。试想有一天，某个人偶然间从位于中亚的里海底部捞上来一颗巨大的钻石、一颗美妙绝伦的红宝石和一颗小小的珍珠，而在这些惊人发现的鼓舞下，寻宝之人纷纷拥入这片世界上最大的湖泊之底，展开了疯狂的打捞，寻找更多的钻石、红宝石、珍珠、翡翠、黄玉和其他宝物，可不管怎么打捞，始终一无所获。人们自然想知道，湖底究竟有没有埋藏着更多的宝石，但是怎么可能知道呢？（附加说明：我的语言略有疏漏，因为我们至少在原则上可以想

象，一支资金充沛的科学考察队伍某一天可能会把整个湖底彻底地打捞一遍，因为湖底虽然很大，却始终是有限的。为了让我的隐喻变得“完全”，我们不得不把里海想象成是没有边界的。读者，把你的想象力再放飞一点点吧！）

接下来，扭转的时刻到了。设想某位具有数学思维的地理学家开始着手证明那两颗璀璨的里海宝石以及那颗小小的圆珍珠都是孤例（sui generis）——换言之，我们有精准的理由相信将不会在里海中发现其他任何形状或大小的珍珠宝石。寻找这样一种证明，说得通吗？怎么可能存在一个天衣无缝的科学理由，能彻底否定在里海湖底发现——除了一颗珍珠、一颗红宝石和一颗钻石以外的——任何宝石的可能性呢？这听起来简直不可理喻。

这正是我们思考物理世界的典型方式——在我们的心目中，它充满了存在变数的事件、有可能翻转的事实以及没有任何根本性的理由可以解释的情况。但是请允许我提醒你，数学家眼中那个纯粹严谨的抽象世界，与我们所有人栖居的那个充满随机和意外的物理世界正好相反。发生在数学世界中的事，在数学家看来，无一例外地都拥有可以陈述且可以理解的理由。

这便是数学家的信条。如果你想理解数学家是如何思考的，你就不得不接受这种思维方式，并拥抱这种思维方式。就这个具体案例而言，斐波那契数列中缺乏整数幂，这虽然在大多数数学家的眼中只是一个小小的谜团，却是格外令人费解的一例，因为它看上去并没有提供任何一条通往答案的自然路径。此中涉及的两种现象——一是指数任意大的整数幂，一是斐波那契数字——似乎在概念上（如里海和宝石一样）彼此相距太过遥远，以至于根本不像有任何深入、系统、不可避免的相互关系。

随后，又有一大批数学家(注)纷纷投入到费马大定理的“大游戏”之中

（这个“臭名昭著”的定理最初由皮埃尔·德·费马于17世纪中叶提出，内容是：没有任何正整数*x*，*y*和*z*，在正整数指数*n*大于2时，能满足$x^n+y^n=z^n$）。这支伟大的国际接力队，把最后一棒交到了安德鲁·怀尔斯（Andrew Wiles）手上，由他发起了伟大胜利的最后冲刺（他的冲刺花了大约8年时间），通过令人惊叹的技术，结合了当代数学宏大地图上的所有观念，终于证明了费马在几个世纪之前提出的论断。

这支队伍的革命性工作打开了全新的路径，似乎把许多闻名遐迩的旧门都撬开了裂缝，其中就包括斐波那契无幂之谜，那扇虽然不大却很吸引人的紧闭之门。在列奥纳多·迪比萨的递归序列中，（除了1以外）在立方数8和平方数144以外再也没有其他的完全幂伙伴了；而大约在费马大定理得到证明的10年之后，三位数学家[注]借助怀尔斯等人的技术，终于得以为其找到了精确的理由。虽然这个理由相当地晦涩艰深，但是隐藏在无限的相互排斥之舞背后的原因，到底还是被发现了。这又是数学家信条的一次伟大胜利——我们又多了一个理由去相信，在数学中，有模式，即有理由。

哥德尔头脑中的一念火花

言归正传，我们继续说哥德尔的故事。他遇见了一个强大有力的观念，即各种各样数不清的数字种类都可以通过不同类型的递归法则获得定义。一种无限的结构或模式从一套有限的初始种子中破土而出、有机生长的意象，对于哥德尔而言已不再仅仅是一种单纯的好奇了；事实上，它让他想起了一个事实，即*PM*中的定理（如同欧几里得《几何原本》中的定理一样）总是（经由推理的形式规则）从*PM*中更早的定理中生发出来的。只有最初的少数几个定理是例外。它们被凭空地强行规定为定理，因而也被称为“公理”（可类比于种子）。

换句话说，那些最初有点模糊的联系在哥德尔的头脑中激发出了一

个精细的类比，*PM*的公理扮演的是斐波那契数列中1和2的角色，而*PM*的推演规则扮演的是把相邻两个数字相加的角色。主要的差别在于，在*PM*中，有不止一条推演规则，所以你每进行一步，都拥有一次选择的机会，不但如此，你还并非必须把选定的规则运用在最近产生的一条或数条定理之上，这甚至又为你提供了更多的选择。但是，除了这些额外的自由度之外，哥德尔的类比是十分贴切的(注)，而最终也带来了相当丰富的结果。

聪明的规则为迟钝的符号灌注意义

我必须在此强调，在*PM*这样的形式系统中，每一条推演规则不仅会把一个或多个输入的公式引向一个输出的公式，而且它们完成这一动作所凭借之物，还是纯粹的排印工具——也就是说，只需要经过纯粹机械化的符号操作，而不要求关于符号意义的任何思考。从一个依照规则制作定理的人（或机器）的视角出发，符号还不如一点儿意义都没有。

另一方面，每条规则都不得不设计得相当仔细，保证当输入的公式表达的是真理时，输出的公式也必须表达一个真理。因此，规则的制定者（这里指罗素和怀特海）就不得不考虑符号的含义，以确保这条规则由不考虑符号含义的操作者（不论是否为人类）操作也能精确有效。

举一个小例子。设符号“∨”意在代表“或”的概念，则一条可能的推演规则为：

从任何一个“P∨Q”公式中都可以推导出反向的“Q∨P”公式。

这条推演规则是合理的，若一个或然陈述（比如“不是你疯了，就是我疯了”）为真，那么反过来的或然陈述（“不是我疯了，就是你疯了”）也为真。

这条特别的翻转v规则，恰好不在*PM*的推演规则之列，但是它本也可以加入进去。要点在于，这条规则表明了人们如何能够在忽略符号意义的情况下机械化地调度符号，同时保存其真理性。这条规则是相当微不足道的，但是还有其他更加微妙的规则，是实实在在地发挥着效用。没错，所谓符号逻辑，本来就是这么一回事儿。它最早由亚里士多德提出，而后又在世纪更迭之际，一砖一瓦地垒建起来[注]，经手的思想家有帕斯卡、莱布尼兹、乔治·布尔（George Boole）、德·摩根（Augustus De Morgan）、弗雷格（Gottlob Frege）、皮亚诺（Giuseppe Peano）、大卫·希尔伯特（David Hilbert）及其他很多人。罗素和怀特海只不过是以一种比他们的前辈更具野心的方式，延续了将推理完全机械化的古老梦想。

把数学家的信条机械化

如果把*PM*的推演规则运用于它的公理（构成“第零代”定理的种子）之上，你就会生产出一些“后代”——“第一代”定理。然后再把这些规则以你能想到的所有不同方式运用在第一代定理（以及公理）之上，你就会生产出一批新的定理——第二代。然后从已得的全部原料之中，又会得出第三代定理，如此继续下去，像滚雪球一样，无穷无尽。*PM*无限巨大的定理体系，完全是由初始的种子和准许从旧定理中推出新定理的排印式“生长规则”所决定的。

不用说，他们希望的是，所有这些机械化生产出来的*PM*定理都是关于数论的真陈述（即不会产生任何假的陈述），反过来说，他们还希望，所有关于数论的真陈述都能作为*PM*定理被机械化地生产出来（即没有任何真陈述是永远不会产生的）。前一种希望，我们称之为<u>一致性</u>，而后一种希望，我们称之为<u>完全性</u>。

简言之，我们想让*PM*定理无限巨大的整体与数论中无限多的真陈

述精确地对应起来——我们想要的是完美无瑕的契合。至少这是罗素和怀特海想要的，而且他们也相信，自己已经凭借*PM*达成了这个目标（毕竟，“$s0 + s0 = ss0$”的确是个定理，不是吗？）。

让我们再来回忆一下数学家的信条。它早在罗素和怀特海出世以前，已经以各种各样的形式存在了几个世纪：

X为真，<u>因为</u>X有证明；

X为真，<u>所以</u>X有证明。

第一行表达的是前面所说的第一种希望：一致性。第二行表达的是前面所说的第二种希望：完全性。我们由此可以看出，数学家的信条与罗素和怀特海的追求具有紧密的联系。然而，他们的目标是把这一信条设定在全新而严谨的基础之上，而*PM*正是支撑它的底座。换句话说，数学家的信条只说到了“证明”，却并没有提及这个词的意义，而罗素和怀特海则想让人们把它的意义理解成*PM*<u>内部的一个证明</u>。

哥德尔本人对于*PM*的力量怀有极大的敬意。他在自己写于1930年的那篇文章的开篇，就表明了这一点：

> 众所周知，随着数学朝着越来越精确的方向发展，它也在很多方面获得了形式化，如此一来，证明便可以根据一些机械化的规则来完成。迄今为止，在已经搭建而成的形式系统中，最为综合全面的，一是《数学原理》的形式系统（PM），一是策梅洛-弗兰克尔的集合论公理系统，后来由约翰·冯·诺依曼（J. v. Neumann）进行了拓展。这两个系统的覆盖范围极为广泛，我们今天在数学中使用的所有证明方法，都已经在系统中完成了形式化，也就是说，它们被简化成了少数的公理及相应的推演规则。

哥德尔向罗素和怀特海的作品给予了慷慨大度的脱帽致敬，可尽管

如此，他却并非真正地相信，真理与*PM*定理之间已经达到了一种完美的契合；他也不相信，这种事情总有一天能够真正达成。而他这种深刻的怀疑来自那座去思维化、机械化、符号化且缺乏意义的数学推理的迷宫式宫殿，他嗅到了一个潜行于其中的极其奇怪的循环。

奇迹般步调划一的同步

在以递归的方式定义的整数序列与交替跃进的*PM*定理集合（就此而言，也可以是任何一种形式系统，只要它拥有发挥种子作用的公理和提供生长机制的推演规则即可）之间，共享着一种相似的概念，这提示哥德尔想到，排印在《数学原理》纸页上的符号模式——也就是从旧定理到新定理的严格逻辑推演——应该可以通过某种方式，在数学世界的内部以一种完全一致的形式“镜照”出来。来自内心的一个声音告诉他，这种联系并不仅仅是一种模糊的相似，而极有可能转化成为一种绝对精确的对应关系。

具体而言，哥德尔想象出了一套自然数，它们像斐波那契的*F*数一样，经由数字运算，从彼此之间有机地生长出来，但同时也能与*PM*定理的集合形成一对一的精准映射。例如，如果你使用排印式的规则*R5*，从定理*X*和*Y*中推导出定理*Z*，与此同时，如果你运用计算的法则，利用数字*x*和*y*算出数字*z*，那么二者之间就是完全匹配的。这也就是说，如果*x*是对应定理*X*的数字而*y*是对应定理*Y*的数字，那么*z*就会“奇迹般地”也成为与定理*Z*对应的数字。完美的同步出现了；（排印的与数字的）两方面的行动步调划一。起初，这种奇迹般的同步只不过是一个小小的思想火花，但是哥德尔迅速意识到，他也许可以让自己朦胧的未尽之梦清晰起来，足以向他人阐明。于是，他便对此展开了顽强不懈的探索。

在公式与大整数之间翻转

为了把自己直觉性的灵感转化成一个严肃、精确而体面的观念，哥德尔首先不得不弄清楚，一串*PM*符号（不论其所言是真还是假，甚或只是一堆随机凑在一处的符号团）如何能够系统地转化为一个正整数，反过来，这样一个正整数又如何能够“解码”还原出它所由来的那一串符号。在这一系统性的映射关系中，每个公式都会有一个数字化的“命名”，这也是如下所示的哥德尔之梦的第一个阶段。

*PM*的原始基础只包括大约10来个符号（其他的符号随后会陆续引入，但它们都是由最初的少数几个符号所定义的，所以从概念上来说，它们并不是必要的），哥德尔为这些符号中的每一个都指派了一个不同的小整数（这些最初的选择都是相当任意的——与某个孤立的符号联系在一起的是哪个数字，其实并不重要）。

遇到多重符号的公式时（顺便说明，在本书中，“符号串”——简称为“串”——和“公式”是同义词），就要从左到右，把符号一个一个地替换成它们的编码数字，然后再把所有那些单个的编码数字组合在一起（方法是把它们按顺序作为连续质数的指数，进行乘方运算），形成一个独一无二的大整数。如此，则一旦孤立的符号被指派了数字，指派给符号串的数字就不再是任意的了。

例如，设（任意）指派给符号“0”的编码数字是2，而符号“=”的编码数字是6，那么对于简单公式“0=0”中的三个符号而言，编码数字就是2，6，2。把这三个数字作为前三个质数（2，3，5）的指数，得到如下数字：

$$2^2 \cdot 3^6 \cdot 5^2 = 72\ 900$$

所以我们看到，72 900就是与公式“0=0”对应的唯一数字。当然，相

比于如此简短的一个公式，这算是一个相当大的整数了。你应该不难想象，与一个由50个符号组成的公式对应的整数，一定是个天文数字，因为我们需要把前50个质数各自取不同次幂，再把这些很大的数字连乘起来，得出一个真正的巨大数字。但是不要紧——不管它们多大，数字也只不过是数字。（质数的个数是无限多的，这对于哥德尔来说是一桩幸事，因为如果打比方说，质数一共只有10亿个，那么他的方法就只能允许他编码由小于10亿个符号组成的公式了。那将会是多大的遗憾呐！）

解码的过程，就是对72 900进行质数因子的分解（这是个唯一解），然后按顺序逐一读出每个质数底数的质数——在这里就是2，6，2。

那么，总结而言，哥德尔以这种并不明显却足够简洁的方式，找到了用一个等价数字替换任何一个给定的*PM*定理的途径。然后，鉴于*PM*中的证明都是公式的序列，他又拓展了“算术化”的观念，令其可以覆盖任意的公式序列，因为他想要拥有处理证明的能力，而非仅仅是孤立的公式。于是，我们也可以凭借本质上完全相同的技术，使用质数和指数，把一个任意长度的公式序列转化为一个很大的整数。你可以想象得到，我们在此谈论的，是真正巨大的数字。

简言之，哥德尔为我们展示出，如何为《数学原理》的诡异记号中任意一种符号模式指派一个独一无二的数字，并使得这个数字可以轻松地解码还原出其所对应的那个视觉模式（即符号序列）。这种精确的双向映射如今被普遍地称为“哥德尔配数”，而对这种映射的构想和打磨，也构成了哥德尔这项工作中关键的第一步。

很大的整数与公式步调一致的移动

接下来的关键一步，就是为特殊整数的集合给定斐波那契式的递归

式定义。通过加减乘除或更复杂的运算，这些特殊的整数可以从先前业已生成的整数中有机地生成出来。一个具有代表性的例子就是*wff*数(注)。这些整数经由哥德尔的算法，代表了“形式良好”或“有意义”的*PM*定理，而不是那些无意义的或不合语法的符号串。又因为*PM*中较长的*wff*都是根据一些简单而标准化的排印并置规则从较短的*wff*中搭建起来的，所以与它们对应的较大的编码数字也同样可以根据简单而标准化的数字运算法则，从较小的编码数字中求得。

上面这些话，我说起来可谓相当之轻描淡写，但事实上，这一步或许已触及了哥德尔核心洞见的最深处——也就是说，一旦符号串完成了“算术化”（也就是被指派了对应的数字），那么任何一种基于规则的对于排印在纸面上的符号串的调度，都可以由某种包含其代理数字的纯粹数字运算给予完美的匹配。其中涉及的数字自然是相当巨大的，但也仍然只不过是数字而已。在罗素和怀特海眼中看起来精妙而复杂的符号调度，在库尔特·哥德尔的眼中看来则是大量的直白的数字运算处理（他自然没有使用“数字运算处理”这个现代术语，因为这些都发生在计算机出现以前的史前时代）。它们只不过是对于同一事件的不同看法而言——这两种看法完全是对等的，可以互相替换。

PM何以扭转至看见自身

哥德尔看到，以递归的方式搭建起一种无穷的数字类别——如*wff*数——的游戏，与斐波那契通过把前面两个数字相加的方式而搭建起*F*数这种数字类别的观念，在本质上是一致的，都是根据某种数字运算处理规则，从更早确定的“俱乐部成员”中制造出新的成员。当然，递归的过程远比把俱乐部中的两个最新成员求和更为复杂。

递归式定义隐含的功能，是在整数的全集中划分出俱乐部的成员与非成员——成员是指那些通过递归式构造早晚可以抵达的数字，而非成

员则是指那些不论经过多久都永远不可抵达的数字。由此可知，34便是*F*俱乐部的成员，而35则是非成员。我们怎么知道35不是一个*F*数呢？这是很简单的——制造新*F*数的规则规定，总是要从更小的数字中制造出更大的数字，所以一旦我们已经路过了一个特定大小的数字，我们就决不再回头，去“拾起”在那附近的其他数字。换句话说，一旦我们已经得出*F*数1，2，3，5，8，13，21，34，55，我们便会知道，它们就是那一段区间内的所有*F*数，所以很明显，34，35，以及一直到54为止，其间的整数都不是*F*数。

因此，若换作另外一种数字俱乐部，根据定义它的递归规则（recursive rules），其输出值有时候比输入值更大，有时候比输入值更小，那么便与*F*俱乐部的简单情况不同，你无法确定自己永远不会折返，去拾起那些在先前路过时遗漏的更小整数。

让我们围绕那个被我们称为“*wff*数”的以递归方式定义的数字俱乐部，再略微做一些思考。我们已经看到，数字72 900是具有“*wff*性”的，但是如果仔细想想的话，你会发现576和2 916都不具备这种性质。（为什么呢？因为，如果你把它们进行因子分解，然后观察2和3的指数，就会看到这些数字所编码的符号串分别是“0=”和“=0”，它们都没有任何意义，也都不是形式良好的公式）。换句话说，如果忽略掉其古怪的定义，*wff*性与平方数性、质数性或者斐波那契的*F*性，并没有高下之分，同样也是纯粹数字世界中一个有效的研究对象。“*wff*俱乐部”的成员与非成员之间的差别，与平方数俱乐部、质数俱乐部或*F*数俱乐部的成员与非成员之间的差别相比，丝毫不差，同样也是一个真正的数论区分，因为*wff*是可以由一种递归的算术（即运算）形式进行定义的。不仅如此，定义了*wff*性的递归规则刚好也总是生产出比输入值更大的输出值，所以*wff*性还与*F*性共享一种简单的特性，即一旦你已经越过了某个值，你就会知道，自己再也不会返回那片值域了。

正如在斐波那契以递归方式定义的序列中存在一个平方数的事实会

点燃一些人的好奇心，同样也会有人对于在*wff*数以递归方式定义的序列中是否存在平方数（或立方数，以及其他次幂）这一问题产生兴趣。他们大可投入大量的时间来钻研这一纯粹的数论问题，完全不必思考其在《数学原理》中对应的公式。

哥德尔的*wff*数本出自罗素和怀特海在《数学原理》中定义良好形式性的规则，但是人们完全可以无视这一事实，这就好比一个人在研究概率的法则时，大可不必在意这门数学的分支是不是源自对于赌博的分析。很久以前启发某个人构想出某种特定的递归式定义之事物，显然对于其所定义的数字没有任何影响；要紧的是，应该有一种纯粹的运算方式，可以通过次数有限的规则运用，从最初的种子中制造出任何一个俱乐部的成员。

如此说来，*wff*数可以相对容易地以递归的方式获得定义，正因如此，*wff*性（同*F*性如出一辙）也刚好成为《数学原理》旨在研究的那类数学概念。可以肯定的是，罗素和怀特海连做梦也没有想过，他们的机械化推理系统竟然还有这样奇妙的用法，把它自身作为机械的性质实实在在地置于了自己的观察之下，简直像是在用一台显微镜检查它自己的镜片有没有破损一样。但世事就是如此难以预料，被发明之物往往会另发明者大吃一惊。

质雅数

哥德尔意识到，沿着怀特海和罗素所著的系列，可以开辟出假设性的一卷，用来定义并系统性地探索*wff*数的不同数字特性，于是他把自己的类比进一步向前推进，启用了很多奇妙的机械装置[注]，但是在概念上并没有遇到太大的困难。他展示出，还有一类更加有趣无数倍的以递归方式定义的自然数，其中的数字对应的是*PM*的可证公式（即定理），我在此将称它们为质雅数[注]（以一种变形的方式向那三卷本著

作的标题致意）。

一个*PM*的证明，当然是从*PM*的公理出发而一路引向目标公式的一系列公式，而且每一步都遵循着某种推理的规则，这些规则在*PM*中成为排印的形式推演规则。针对作用于*PM*串的每一条排印的推演规则，哥德尔都给出了一条与其完美匹配的作用于数字的运算法则。数字的运算对排印式的操作嗤之以鼻，还粗鲁无礼地说，“凡是你能干的，我都干得比你更好！”[注]好吧，并非真的更好——但关键的要点在于，哥德尔清楚无疑地呈现出，运算法则总是能够，以绝对一致的步调完美地模仿任何一条排印的形式规则，所以算术法则就是刚刚一样好。

结果就是，在罗素和怀特海的形式系统中，每一个可证串都有一个相应的质雅数。每一个属于质雅数的整数都可以解码出符号，而你得到的符号串便是一个在*PM*中可证的公式。同样，任何一个在*PM*中可证的公式都可以编码成为一个异常庞大的整数，而且上帝作证，只要经过足够多的计算，你总可以得出，这个数字是一个质雅数。我们要举一个质雅数的简单例子，还可以找我们的老朋友72 900，因为“0=0”这个公式不仅仅是一个形式良好的公式，还是一个可以在*PM*中推演出来的公式，而这并不怎么令人感到惊讶。（毕竟，如果它不能在*PM*中推演出来，那么*PM*作为一个关于数学推理的机械模型就真的太可悲了！）

*PM*的推演规则有时候会产生比输入串更短的输出串，这一事实造成了*wff*数与质雅数之间的一个关键性差别。这就意味着，定义质雅数的相应算术法则，有时会允许在输入一个大的质雅数后，输出的是一个更小的质雅数。因此，被访问过的一段数字队列，总是有可能在后来再次被访问，而因为这一事实的存在，判断一个给定的整数是否属于质雅数，就变得难上加难了。对于质雅数而言，这是一个位于核心且相当深刻的事实。

与平方数、质数、*F*数和*wff*数遭遇的情况一样，沿着怀特海与罗素

著作系列，还可以在假想中开辟出另一卷，为质雅数赋予定义，并研究它们的数学特征。例如，这一卷中也许包括一个对于（在经过仔细考察后）声明了“72 900是质雅数”的*PM*公式的证明，也许还包括关于另一个呈现出相反声明（“72 900不是质雅数”）的公式的讨论。当然，后一种陈述是假的，而前一种陈述为真。

使用*PM*的记号，甚至还能表达出更加复杂的数论观念，并在假想的一卷中进行讨论，比如“质雅数有无穷多个”——这完全等价于（经由编码）“在*PM*中有无穷多个可证的公式”这一声明。

人们还可以提出18世纪风格的数论问题，如“哪些整数可以表达成两个质雅数的和，而哪些整数不能？”虽然这看上去有点古怪，而且很可能没人会认真地提出这样一个乖僻的问题，但是这里要说的重点在于，即便质雅数的特性是一种相当晦涩艰深的“现代”特性，但它也是一个整数的数论特性，与诸如平方数、质数或斐波那契数的“经典”特性一样货真价实。

质雅数的神奇力量

设想有个人告诉你，他们制造出了这样一台机器——我给它起个绰号，叫“格鲁大师”，它面对以“n是一个质数吗？”为形式的问题，总能给出正确的答案，把n换成你想问的任何一个整数都可以。当被问到“641是一个质数吗？”这个问题时，格鲁大师会转动一下它的轮子，然后说：“是。”而被问到642时，格鲁大师会“思考”一下，然后说：“不是。”我猜，这样一台机器并不会让你觉得惊为天人。时至今日，这样一台机器是可以实现的，不论其使用的是硅晶电路还是多米诺链的技术，都不会让人惊讶。

但是，再设想有人告诉你，他们制造出了一台类比机器——我给它

起名叫“格鲁大师”，它总能正确回答任何以“n是一个质雅数吗？”的形式提出的问题。这份声明与上面那个声明具有严格的类比关系，而它给你的印象，还是同样地索然无味吗？如果真是这样，我只能诚恳敬重地说，你的看法很快就要改变了。

理由如下。如果你相信格鲁大师是可靠的，同时也信仰数学家的信条（《数学原理》版），那么你就能推断出，你这位小小的格鲁大师，可以凭一己之力，回答出你感兴趣的任何一个数论问题，简直就像从神灯里召唤出的精灵一样。怎么会这样呢？是什么把格鲁大师变成了魔法强大的精灵呢？

好吧，假设你想知道，陈述X是真的还是假的[注]（例如，“每一个比2大的偶数都是两个质数之和”这个著名论断[注]——正如我在前文所说，在经过了近3个世纪的努力之后，它至今仍悬而未解）。你只需要把陈述X用*PM*的形式符号写下来，然后把得到的公式转换成它的哥德尔配数x，再把那个数字递给格鲁大师（由此询问x是否为一个质雅数）。当然了，x将会是一个巨大的整数，所以格鲁大师可能需要经过很长一段时间之后，才能给你答案，但是（假定格鲁大师不是一个江湖骗子）它早晚都会吐一个“是”或“不是”出来。如果格鲁大师说了“是”，那么你就会知道，x是一个质雅数，这告诉你，x编码的公式是一个可证的公式，而这也就意味着，陈述X是真的。反之，若是格鲁大师告诉你“不是”，那么你就会知道，陈述X是不可证的，所以，相信数学家信条（《数学原理》版）的你，就会得出其为假的结论。

换句话说，但凡我们能拥有一台可以屡试不爽地区分出质雅数与“粗俗数”（non-prim）的机器，并同时默认《数学原理》版的数学家信条是有效成立的，我们就能分毫不差地区分出所有的真陈述与假陈述。简言之，格鲁大师将赐予我们解开一切数学知识的金钥匙。

由此看来，质雅数如一领披风，包含了卷在其中的一切数学知识！

在哥德尔之前，没有任何人曾梦想过任何拥有这种充满魔力的神谕性的其他数字序列。这些神奇的数字简直是无价之宝！但是我也告诉过你，质雅数是捉摸不定的，因为小的数字有时会在后面的阶段才加入俱乐部当中，所以想要区分出质雅数和粗俗数，或者说打造一位格鲁大师，都绝非易事。（此处意在对后面的内容进行一次预言。）

哥德尔式的怪

最终，哥德尔把他的类比引向了一个难以绕过的重大结论，即为他的读者拼写出了（当然不是一个符号一个符号地写下来的，而是通过一套精确的"汇编指令"）一个具有天文数字级长度的*PM*公式，而这个公式只做出了一个看似相当无害的声明："某个特定的整数*g*不是质雅数。"然而，在这个公式中所说的那个"特定的整数*g*"，刚好以一种绝非偶然的（有人会说是恶魔式的）巧合，取了与这一个公式联系在一起（即作为其编码）的数字（因而它必定是一个极其庞大的数字）。正如我们即将所见，哥德尔的古怪公式可以在两个不同的层面上进行解读，而根据人们解读方式的不同，也含有两种十分不同的意义。

在其最直白的层面上，哥德尔的公式仅仅是声明了，这个庞大的整数*g*不具备人称质雅数性的数论特性。这个声明与"72 900不是一个质数"很像，只不过*g*肯定比72 900要大很多，而质雅数性也是比质数性远为更加棘手的特性。不管怎么说，根据哥德尔对质雅数性的定义，它是在*PM*系统规则下的符号串可证性的数字镜像，所以这个公式同时声明了：

这个编码数字刚好为g的公式，在《数学原理》的规则下是不可证的。

我已经说过，这个编码数字"刚好"为*g*的公式，正是做出如上声明的公式。简言之，哥德尔的公式做出了一个关于自己的声明——即如下

声明：

这个公式本身在PM的规则下是不可证的。

有时候，这第二种措辞又被直白地表达为“我不是一个定理”，或者甚至更加简洁（默认是“在*PM*系统中”）：

我不可证。

哥德尔进一步展示出，他的公式虽然非常奇怪，而且第一眼看上去让人感到头晕眼花，但其实并没有那么不常见；实际上，它只不过是就*PM*系统做出各种声明的无限大的公式家族中的一员而已，还有很多其他的成员也做出了同样怪异和扭曲的（有真有假的）关于自己的论断（例如，“我和我的否定命题都不是*PM*的定理”，“如果我在*PM*内部有一个证明，那么我的否定命题便有一个比我更短的证明”，不一而足）。

罗素和怀特海在他们的三卷本鸿篇巨制《数学原理》里，定义了一个朴素严谨的形式化世界，由于得到了类型论的保护，这个世界本该不受任何悖论的侵扰，可是年轻的库尔特·哥德尔—1931年的时候，他才25岁——却发现了由万万料想不到的、出奇扭曲的公式汇聚成的一片海洋。自哥德尔最初提出那些公式以来，不计其数的近亲公式如雨后春笋般涌现，而它们所具有的大量违背直觉的特性，令数学家、逻辑学家和哲学家们魂牵至今。

如何把一个公式的哥德尔配数插入到公式之中

在结束有关哥德尔这项伟大壮举的话题之前，我不能不先讨论一个略微有点技术化的问题，因为如果我漏过了这个问题，有些读者肯定会对哥德尔工作中的一个关键面向困惑不解，甚至产生怀疑。更别说，这

本身就是一个相当不可思议的观念，所以很值得在此简略一提。

这个让人不安的问题是这样的：哥德尔到底是如何把一个公式的哥德尔配数纳入这个公式本身的？你刚开始思考这个问题的时候，它看起来就像要努力把一头大象塞进一个火柴盒里似的——而且从某种意义上来说可不就是这样。没有任何公式能够在字面意义上包含㊟与它自己的哥德尔配数对应的数词，因为这个数词含有比这个公式多得多的符号！最初看起来，这似乎是一块注定迈不过去的绊脚石，但事实却并非如此——如果你回想一下我们有关贝里悖论的讨论，也许就能明白其中的关键了。

这里的要诀有赖于一个简单的事实，即某些巨大的数字具有很短的描述（例如，387 420 489，用五个字就可以描述出来：9的9次方）。如果你能用一个简短的算式计算出一个很长的公式的哥德尔配数，那么你就不必再以笨重而臃肿的方式去描述那个数字（“0的后继数的后继数的……”），而是可以通过你的运算式捷径来描述它；而如果你用以表达这个捷径的工具是公式内部的符号（而不是插入的数词本身），那么你就可以让这个公式自己谈论自己，并且不必花费把大象塞进火柴盒的吃奶力气。我不会企图用数学的方法对此进行解释，但是我会给出一个优雅的语言学上的类比㊟，代以阐明其要旨。这个类比借用自哲学家W. V. O.奎因（W. V. O. Quine）。

哥德尔的火柴盒塞大象要诀对阵奎因的类比

设想你要写出一个汉语句子，令其在不使用“本句”这个短语的前提下谈论自身。你将会发现，这个挑战是相当棘手的，因为你不得不在这句话的内部，使用引号和引语，切实地描述出这句话。例如，先看看下面这个（多少有点白费力气的）初次尝试：

“本句话有7个字”这句话有7个字。

虽然我刚刚写出（而你刚刚读到）的这句话为真，但遗憾的是，它并没有谈论其自身。说到底，上面这个完整的句子包括**14**个字，还外加一对引号。这句话谈论的是通过引号嵌于其中的一个更短的句子。而把“7”改成“14”，仍不会让它反指自身；这个简单的愚蠢动作，只会把我那个原本为真的句子变成假的而已。可以看一下：

“本句话有14个字”这句话有14个字。

这句话是假的。更重要的是，它*依然*还是在谈论一个嵌于其中的更短的句子。如你所见，我们目前还没有接近设计出一个谈论其自身的句子。

问题是，不管我在引号里放的是什么，它必然比其所嵌入的整句话更短。这是一个显而易见的事实，也构成了“把一个公式的哥德尔配数直接插入其自身”这块绊脚石在语言学上的一个精确类比。一头大象是塞不进一个火柴盒的！可另一方面，一头大象的DNA却很容易*就能*放入火柴盒内……

确实，除了大象本身之外，DNA也可以描述一头大象，类似地，想要绕过这块绊脚石，我们也可以使用这个巨大数字的*描述*，而不是数字本身。（说得更准确一点，我们可以用一个简洁的符号描述来代替一个巨大的*数词*。）哥德尔发现了这个要诀，虽然它相当微妙难懂，奎因的类比却让它变得十分容易理解。请看下面这个句子片段，我将其称为“奎因的准妙语”：

如果紧跟在置于引号中的自己之后，可以造出一个完整的句子。

如你所见，奎因的准妙语显然*不是*一个完整的句子，因为它在语法上缺少一个主语（即“造”的主语是缺失的）；这也是我为什么要在它前

面加上一个“准”字的原因。但是，如果我们在这句准妙语的开头加一个名词——比如，“奎因教授”这个头衔——又会怎样呢？这样一来，奎因的准妙语就会变成一个完整的句子，我也可以叫它“奎因妙语”了：

“奎因教授”如果紧跟在置于引号中的自己之后，可以造出一个完整的句子。

在这里，动词“造”确实拥有一个主语了——即奎因教授的头衔，受到一个由15个字组成的条件从句的限制。

但是，奎因妙语有什么意义呢？为了弄清楚它的意思，我们不得不真正构造出一个它所谈论的实体，这就意味着我们不得不让奎因教授这一头衔紧跟上放在引号中的自身。我们由此得到：

“奎因教授”奎因教授

我们刚刚创造出的奎因妙语只是声明（或者说断言）了这个有点傻的短语是一个完整的句子。好吧，这一论断显然是假的。上面那个短语不是一个完整的句子；它甚至连一个动词都没有。

话说回来，我们使用奎因教授的头衔是一个相当任性的选择，此外还有上百万种选项。有没有某个其他的名词，我们可以把它放在奎因的准妙语的开头，令奎因妙语为真呢？由哥德尔所意识到，且经由奎因的类比而清晰表明的道理是，要想做到上述这一点，你不得不使用一个无主语的句子片段作为动词“造”的主语。

这种无主语的句子片段，有可举的例子吗？很简单，随便拈来一个常见的句子，比如“雪是白的”，然后砍掉它的主语就好。你由此便可得到一个无主语的句子片段：“是白的”。让我们把这个片段作为一个名词，放在奎因的准妙语的开头：

“是白的”如果紧跟在置于引号中的自己之后，可以造出一个完整的

句子。

这句不长不短的拗口的话做出了关于某种建构的声明，而这种建构尚有待我们呈现，所以让我们闲话少说，直入正题：

“是白的”是白的。

（我额外多加了一个句号，但是让我们就别在这里吹毛求疵了。）

现在，我们刚刚所造的显然是一个完整的句子，因为它有一个动词（“是”），而且这个动词还有一个主语（加引号的短语），这全都说得通。请注意，我并没有说它是真的，因为它明白无疑是假的：“是白的”事实上是黑色的（虽然，公平地说，汉字及笔画的空白处确实包括一些白色的空间，不然我们是无法阅读它们的）。不管怎么说，当把“是白的”输入给奎因的准妙语后，确实造出了一个完整的句子，而这正是奎因妙语所声明之义。我们切切实实地在向前推进。

最为神奇的一步

我们最后一个邪恶的把戏，是把奎因的准妙语本身作为一个名词放在它的开头。如此一来，奎因的准妙语前面，便加上了一个放在引号中的自己的复制品：

“如果紧跟在置于引号中的自己之后，可以造出一个完整的句子”

如果紧跟在置于引号中的自己之后，可以造出一个完整的句子。

这个妙语声明了什么呢？我们首先必须确定，它在进行关于什么实体的谈论，这也就意味着，我们必须构建出可类比于“‘是白的’是白的”之物。这个类比物如下：

“如果紧跟在置于引号中的自己之后，可以造出一个完整的句子”

如果紧跟在置于引号中的自己之后，可以造出一个完整的句子。

我希望你到此仍能跟上我的脚步，因为我们已经触及了问题的核心。奎因妙语谈论的是一个与奎因妙语完全一致的短语！它断言某句话是一个完整的句子，而当你建构起那句话之后，会发现它就是奎因妙语本身。所以，奎因妙语谈论的是它自己，并断言它自己是一个完整的句子（它当然是个完整的句子，尽管它是由两个无主语的句子片段拼接组成的，一个在引号内，一个在引号外）。

你大可以在此多琢磨一会儿，而我就要跳回一切的源头了，也就是哥德尔提出的谈论自身的*PM*公式。要点是，既然哥德尔配数可以作为公式的名字而插入到公式之中，那么它们就完全可以类比于放在括号中的短语。而正如你刚才所见，我们有办法使用引号和句子片段造出一个谈论其自身的句子（或者如果你愿意的话，也可以说它谈论的是另一个句子，只不过这个句子是原句的克隆体，所以两个句子在任何方面都是一致的）。

哥德尔类似地创造了一个“无主语的公式片段”（我的意思是，这个*PM*公式并非针对某个特定的整数，其处理的是某个未知的变量x）。然后，他又进行了相当于把奎因的准妙语（加上引号）递送给它自己的操作，用公式片段的哥德尔配数k（这是一个具体的数字，不是一个变量）替换了变量x，由此产生出一个对更大的整数g做出论断的公式（而不仅是一个片段）。而g又是这一论断的哥德尔配数。最后不容忽视的一点是，这个论断并非关于所言实体是否为一个完整的句子，而是关于所言实体是否为一个可证的公式。

火柴盒中的大象既不是鱼，也不是鸡

我知道，想要一口气把我说的这些彻底搞明白是有点困难的，所以如果你一时不能理解（事实上，的确需要仔细重读很多遍才能真正理解），请千万不要泄气。曾有很多成熟的数学家向我承认，他们从来都没有真正完全理解过这些论述！

我想，为了有助于更好的理解，我应当在此关键之处再展示一种杂交句，用奎因的语言传达出哥德尔自指建构的精髓——也就是使用我们刚刚讨论的那种格式。这种杂交句是长成这个样子的：

“当输入其自身的哥德尔配数时，会得出一个非质雅数”

当输入其自身的哥德尔配数时，会得出一个非质雅数。

上面这个句子既不是鱼，也不是鸡，因为它不是一个属于《数学原理》的公式，而是一个汉语句子，所以它自然也不具有哥德尔配数，也不可能是一个*PM*定理（或者非定理）。这是多么混杂的隐喻！

可即便它是如此混杂的一个隐喻，仍然出色地完成任务，传达出了真正由哥德尔配数构造的*PM*公式的神韵。你只须时刻记住，使用引号是表示取哥德尔配数的一个隐喻，所以上面一行应该被想象成一个哥德尔配数（k），而不是一个引号中的句子片段。这就意味着，在隐喻的层面上，下面一行（一个汉语句子片段）<u>已经</u>在用它自己的哥德尔配数作为自己的主语了。真是有趣！

我知道这不太好理解，所以让我换一种些许不同的方式，再陈述一遍。哥德尔要求你想象由k代表的公式（那个公式刚好涉及变量x），然后把k输入进去（这就意味着把单个字母x换成极长的数字k，这会让你的公式变得比一开始庞大很多），之后再取所得结果的哥德尔配数。那个数字是g，远远比k巨大——最后，哥德尔宣称，<u>这个</u>极大的数字不是一个质雅数。如果你跟上了我这番避实就虚的论证，你就应该同意，这个哥德尔配数（g）并没有全盘呈现在公式内部，而是由这个公式十分

巧妙地描述了出来。大象也是通过DNA的描述才被装进了火柴盒中。

斯鲁戈和“莫顿盐”女孩

好吧，我不想在此强调技术细节。我们需要记住的重点是，哥德尔设计出了一个非常巧妙的数字描述要诀——从一个不那么巨大的数字k得出十分巨大的数字g的配方，由此得到了一个声明它自己的哥德尔配数为非质雅数的*PM*公式（这就意味着，这个公式实际上声明了它自己的非定理身份）。同时，你还应当努力记住，那个“小”数字k是一个包含变量x的“公式片段”的哥德尔配数，可以类比于放入引号当中的无主语的句子片段，而更大的数字g则是与一个以*PM*记号表达的完整句子对应的哥德尔配数，可类比于一个完整的汉语句子。

这种自我指涉的乐趣与通俗文化绝非格格不入，事实上，我们在此进行对比的两个观念——一个直接包含其自己的哥德尔配数的公式（它将必然引发一个无穷的回归）与一个包含了对其哥德尔配数的描述的公式（优美地绕过了无穷的回归）——在读者们也许很熟悉的两幅图画中得到了迷人的阐示。



在第一幅图画中，厄尔尼·布什米勒（Ernie Bushmiller）的经典连环画《南茜》（*Nancy*）中的人物斯鲁戈梦到自己在梦到自己梦到自己梦到自己，如此无穷无尽。这显然是一种自我指涉的情况，但是它还涉及了一种无穷的回归，可类比于一个直接包含其自身哥德尔配数的*PM*公式。遗憾的是，这样一种公式只能无限地延长下去！

相比之下，我们的第二幅图像，即莫顿盐盒子的著名商标，展现的是一位抱着莫顿盐盒子的女孩。你可能以为自己又嗅到了无穷回归的味道，但若真是这样的话，你就上当了！女孩的胳膊遮住了本应发生回归的关键部位。如果你想请求这个女孩把她的盒子递给你，好让你切实地看到发生在商标上的无穷回归，那么你最终还是会失望而归，因为在那只盒子上的商标里，还会有一个小女孩用胳膊挡住一只更小的盒子上回归发生的部位。



可尽管如此，我们还是得到了一幅自我指涉的画，因为食品杂货店里的消费者都能理解，这个商标上展现的盒子和他们手中拿的那个盒子是一样的。他们是怎么得出这个结论的呢？是通过类比的运用。具体地说，他们不仅拥有拿在自己手中的大盒子，还能看到女孩抱着的那只小盒子，而这两只盒子之间存在着大量的共同点（圆柱的形状、深蓝的颜色、两边的白色顶盖）；如果这些尚不够用，他们还能看到小盒子里撒出的盐。这些证据组合在一起，足以说服每个人相信，小盒子和大盒子是完全一致的，于是你也就得到了想要的东西：一个不含无穷回归的自我指涉！

在本章的结尾，我想明确地点出，用汉语翻译哥德尔的公式及其近亲时，最简洁的一种方式启用了“我”这个词（“我在*PM*中不可证”；“我不是一个*PM*定理”）。这并非巧合。事实上，在哥德尔那个朴素严谨的

数学怪圈与人类对于一个有意识的自我的概念之间，存在着一种深刻的联系，而正是这次对于“我”的颇不正式且几乎懒散的使用，才让我们得以首次一窥其中的关联。



---

1. See [Huntley].

2. See [Péter] and [Hennie].

3. A recent book that purports to convey the crux of the elusive ideas of this team is [Ash and Gross]. I admire their chutzpah in trying to communicate these ideas to a wide public, but I suspect it is an impossible task.

4. 他们是雅恩·卜高德（Yann Bugeaud）、莫里斯·米尼奥特（Maurice Mignotte）和萨米尔·西科塞克（Samir Siksek）。结果，证明144是斐波那契数列中（除了1之外）唯一的质

数其实并不需要高度抽象的观念，尽管还是十分微妙的。这个证明在1964年由约翰·科恩（John Cohn）完成。

5. The essence and the meaning of Gödel’s work are well presented in many books, including [Nagel and Newman], [DeLong], [Smullyan 1961], [Jeffrey], [Boolos and Jeffrey], [Goodstein], [Goldstein], [Smullyan 1978], [Smullyan 1992], [Wilder], [Kneebone], [Wolf], [Shanker], and [Hofstadter 1979].

6. See [Nagel and Newman], [Wilder],[Kneebone], [Wolf], [DeLong], [Goodstein], [Jeffrey], and [Boolos and Jeffrey].

7. “*wff*数”是形式良好的公式的简称，“0+0=*sss*0”即为一个典例。虽然它的命题为假，但仍然是一个有意义的陈述。相形之下，“=）0（=”和“00==0+=”就不是*wff*了。就像我们随便举出的一个假词串“zzip dubbiwubbi pizz”那样，它们什么意思都没有表达。——译者注

8. 这一术语指程序或者算法。——编者注

9. 质雅数（prim number），是假想中与*PM*形式系统对应的一种数字，每个质雅数对应*PM*系统中的一个可证公式。作者对这类数字的命名，运用了双关手法，一方面与质数（prime number）类比，一方面借用英文单词prim（雅致）的意义，暗示这类数的特征。为了与图论中的“prim算法”做出区分，在译文中不保留prim原文，而是译作“质雅数”。——译者注

10. 我亲爱的朋友丹尼尔·丹尼特曾经写过一篇有趣的书评，其中有这样一句话：“‘凡是你能干的，我都能在更高的层面上做’（anything you can do I can do meta）是道格的一句座右铭，而且显然他在自己做的每一件事上反反复复地实践了这句话。”好吧，丹的这句古怪的话让人以为是道格自己想出了这句“座右铭”，并且真的在到处宣扬。（要不然丹为什么要把它放到引号里？）事实上，我从来没有说过这样的话，也没有这样的想法，而丹就是用他专属的方式“给了我一个更高的层面”（going me one meta）。但是，让我惊讶的是，这句“座右铭”开始在人们之间传播开来，并回过头来引用给我听，仿佛我真的想到过这句话，而且真心相信。虽然丹的这句座右铭既聪明又有趣，但是我很快就觉得厌烦了，因为它根本就不符合我的自我想象。不管怎么说，这个注释只是我为了掐断这个谣言而做的一次小小尝试。上面引号里那句话不是侯世达说的话。但是我猜我的这次尝试基本上不会有什么效果。

11. The dream of a mechanical method for reliably placing statements into two bins——“true” and “false”——is known as the quest for a *decision procedure*. The absolute nonexistence of a decision procedure for truth (or for provability) is discussed in [DeLong], [Boolos and Jeffrey], [Jeffrey], [Hennie], [Davis 1965], [Wolf], and [Hofstadter 1979].

12. 即哥德巴赫猜想。——编者注

13. [Nagel and Newman] presents this idea very clearly, as does [Smullyan 1961]. See also [Hofstadter 1982].

14. See [Quine] for the original idea (which is actually a variation of Gödel's idea (which is itself a variation of an idea of Jules Richard (which is a variation of an idea of Georg Cantor (which is a variation of an idea of Euclid (with help from Epimenides))))), and [Hofstadter 1979] for a variation on Quine's theme.

第11章

类比如何创造意义

PM公式的双重相关性

《数学原理》是一座令人望而生畏的智力要塞，建造者为了抵抗可怕的自我指涉之灾乱，煞费苦心地把它打造成为一座碉堡；而年轻的奥地利“顽童”库尔特却白纸黑字地宣称，实际上，它已经千疮百孔，被那些据说陈述着各种各样关于自己的荒谬而不可理喻之事的公式打成了筛子。想象一下，新近受封的罗素公爵会有多么困惑不解。怎么可以容忍这样一种骇人听闻的暴行发生呢？那些空洞琐碎而喋喋不休的自指命题，怎么可能穿过分支类型论（Theory of Ramified Types）这座美丽而永恒的厚厚的城墙，偷偷潜入进来呢？这个自命不凡的奥地利巫师肯定是施展了某种邪恶的咒语，可他到底是通过什么手段作恶的呢？

答案就是，哥德尔在他那篇经典的文章——《论<数学原理>及其相关系统中的形式不定性命题（I）》[注]——中，重新分析了意义的概念，并总结出，一个*PM*的公式所包含的意义，并没有罗素以为的那样简单——那样确定无疑。公道地说，罗素本人确实总是坚持认为，*PM*中那些看起来很奇怪的长公式没有任何内在的意义。的确，鉴于*PM*的定理是从毫不重视意义的形式规则中批量生产出来的，罗素常说，整套东西只不过是一队无意义的记号列阵而已（正如你在第9章结尾处所见，《数学原理》中的页面，常常看起来更像是某种其他领域的艺术品，而不是数学著作）。

尽管如此，罗素还是不忘小心翼翼地指出，所有这些形状奇特的模

式，不论是马蹄形、钩子形、星形还是歪歪扭扭的蠕虫形，只要人们愿意，都可以被解读成关于数字及其属性的陈述，因为在压力的迫使之下，一个人可以把无意义的竖立的蛋形“0”认作数字零，而把同样无意义的交叉符号“+”认作是加号的代表，依此类推，则*PM*中的所有定理都可以变成关于数字的陈述——但并不仅仅是胡说八道。只要想象一下，如果蠕虫形状的“ss0 + ss0 = ssss0”也能成为一个*PM*的定理，罗素该有多么崩溃！在他而言，这简直是最为重大的灾难了。正因如此，他才不得不退一步承认说，在他那部晦涩难解的卷帙中，存在可以被发现的意义，（不然的话，他为什么要花费人生中那么长的时间来写作它，又为什么会在意，哪一串符号才是真的定理呢？）但是要想找到那种意义，依赖于使用一种映射关系，把纸页上的形状与抽象的数值（例如，0，1，2……）、运算（例如，加法）、关系（例如，相等）、逻辑概念（例如，“非”“与”“有”“全”）等等联系起来。

罗素在解读他那座符号堡垒的意义时，对于一种系统化映射的依赖是相当显而易见的，因为年轻的“顽童”哥德尔所发现的，也仅仅是一种不同的系统化映射（必须承认，这是一种更复杂的映射），令人得以从那同一座堡垒中读出不同的意义。这么说来，颇具反讽意味的是，哥德尔的发现在很大程度上遵循着罗素的精神。

哥德尔的精妙的新编码，在符号串与数字之间建立了系统化的相互映射（也不要忘了，它还在排印的调度法则和数字运算之间建立了双向映射），得益于此，很多公式都可以在第二种层面上获得解读。经由旧的标准映射而得到的第一层意义，恰如罗素所称，总是与数字相关，而使用哥德尔（踩在罗素的第一种映射肩上）后来提出的映射得到的第二个层面的意义，则与公式相关。既然两个层面的意义都依赖于映射，那么哥德尔的新层面的意义与罗素的原始意义相比，无论是在真实性还是有效性上，都丝毫不差——只不过多少有点儿更难看见而已。

免费添加格外的意义，多亏你了，类比！

我曾用很多年的时间来思索哥德尔在1931年所做之事，而令我最为震惊的是他投向意义根底的洞见——他发现，借助于一种映射，成熟丰满的意义可以在完全出乎意料的地方忽然现身。我发觉，这份洞见极其深奥，又分外简单。可是奇怪的是，我几乎从未看到过有人在谈论这个观念时曾触及我所发现的深奥性，所以，我决心亲自出马，在本章中迎接这份挑战。为了达到这个目的，我将使用一系列由相当平凡琐碎变得逐渐愈加精妙的范例，希望它们同时也能一个比一个好玩儿。下面就让我们开始吧。

我同一位朋友在咖啡馆里排队，看到柜台后面的一个大浅盘里有一块巨大的巧克力蛋糕，于是我让服务员帮我切一片。我的朋友也蠢蠢欲动，但是抵制住了诱惑。我们来到桌子前，我咬了一口蛋糕，然后说："哎呀，真难吃！"显然，我说这话的意思是，不仅我拿到的这一块很糟糕，而且整个蛋糕都不怎么样。这样一来，我的朋友应该为自己抵制住了诱惑而感到明智（或幸运）。这句平常的评价，演示了我们如何能够毫不费力地向外进行泛化。我们无意识地想到，"这一块蛋糕和整块蛋糕的其他部分很像，所以关于它的评价也能同等地应用在其他任何切块之上。"（这里还存在一个假定的类比，就是我的朋友对食物的反应与我相似，但是这一点我就先不管了。）

让我们再试试另一个例子，微微放开一点手脚。在派对上的一个盘子里有一些曲奇，我拿起一个，咬了一口，然后对我的孩子们说："真好吃！"我的孩子们立马每个人抢了一块。为什么？因为他们想尝尝好吃的东西。这没错，可他们是怎么从我对*我的*曲奇做出的评价里一跃而得出有关盘子里*其他*曲奇的结论的呢？答案很明显，这些曲奇在某种意义上都是"一样"的。但是，与一块蛋糕的多个切块不同，这些曲奇并不是某个单一物体的组成部分，因而它们同蛋糕切块相比，彼此之间"更加不同"一点点——但是它们都是由同一个人用同样的原料和相同的厨

具做出来的。这些曲奇都出自同一批——它们属于同一类。从各个方面来说，我们都认为它们是可以互相替换的。当然，每块曲奇都是独一无二的，可是从人类进食曲奇的角度来看，它们几乎必然是等价的。因此，如果我对某块特定的曲奇说："天啊，真好吃！"那么我这句陈述的含义就会在类比的动力作用下，自动传递给每一块曲奇。不消说，当所有的曲奇都装在同一个盘子里时，由一块曲奇推断另一块曲奇，这自然是一个相当微不足道的类比，可尽管如此，它终归还是一个类比，而且它令我的"真好吃！"这句特定陈述得以立即被当作了关于整盘曲奇的普遍陈述。

你也许觉得这些例子说起来太幼稚了。第一个例子包含了一个在同一块蛋糕的不同切块之间的"类比"，而第二例子则包含了一个在同一个盘子里的几块曲奇之间的"类比"。这些乏味无奇的滥调真的配得上"类比"这个标签吗？在我看来，这是毋庸置疑的；事实上，正是无数个为人忽视并随手丢弃的类比，构成了我们丰富的精神生活的绝大部分的厚密质地，而它们并不比这两个例子有何特殊之处。可我们把这些随手丢弃的类比视为理所当然之物，并倾向于认为，"类比"这个词必须搭配给某种远为超脱之物。但是，在我的生命中最常反复出现的一首主题曲就是，我们应当对看起来最平常不过的类比给予极大的尊重(注)，因为如果对它们详加审视，就会发现，它们通常都萌生自人类认知的最深处，并揭示了其最深处的根底。

在日常情景中利用类比

如上所见，一则旨在讨论情境A的评论，也可以隐含地适用于情境B，即便原本没有任何讨论B的意图，甚至于对B完全未曾提及。只需要一个简单的类比，即一种揭示出两种情境具有本质上一致的核心结构或概念内核的轻松映射，就可以制造出额外的意义以供解读，而至于解读

与否，则是个人的选择。简而言之，有关一种情境的陈述可以听起来仿佛也与一种类比的情境有关——或者，用稍微专业一点的术语说，是一种同构的情境。同构性就是一种形式化的、严格的类比——把两种情境之间的相似性网络明确而精准地表述出来，而我在下文中会经常用到这一术语。

如果情境A与情境B之间的一个类比明显到了直白的地步（不管它有多么简单），我们有时候就会利用这个类比，“假装不经意”地通过表面上单纯谈论A而去谈论B。“嘿，听着，安迪——进屋之前，把你的泥靴子脱掉！”这句话是喊给五岁大的儿子安迪听的，可是和安迪一起踩在门口的小朋友比尔，靴子也一样渗着泥水，所以这句话显然也是说给比尔听的，只是借助了一个简单而显见的类比（你要是愿意的话，可以把这称为从男孩到男孩的跳跃，很像之前从曲奇到曲奇的跳跃）。通过类比进行暗示，使我们得以不失礼貌而又有效地传达出信息。当然，我们须得相当确定，我们把隐含的信息传递给的那个人（这里是比尔）很可能明白A和B之间的类比，不然的话，我们这个机智圆滑的手腕就完全没用了。

让我们沿着例子的链条继续向上攀登。身处浪漫情境中的人们总是会使用这样的伎俩。在某个夜晚，泽克西斯与尤兰达温柔相拥，当激情正浓的时刻，泽克西斯问他的宝贝甜心：“我有口臭吗？”他是真心地想知道答案，这在他是非常贴心的，但与此同时，他的问题还（不论是有意还是无意地）负载着另一个层面的意义，而这就没那么贴心了：“你有口臭！”尤兰达回答了他的问题，但是她自然也在一瞬间领会了其中隐含的另一层意味。事实上，她怀疑泽克西斯的真实意图就是要把她的口臭告诉她，而不是询问他自己的情况——他只是在用一种委婉的方式而已。

那么，一句陈述如何能在同时表达出两种层面的意义？第二个意义如何能够潜行在第一个意义之中？亲爱的读者，你和我都清楚答案是什

么，可还是让我把它挑明了吧。同泥靴子的情境一样，在这两个爱人之间也存在着一种非常简单、非常显著、非常鲜明、非常明显的类比，这也就意味着，任何一句有关X的陈述都将（或者至少可以）被同时视为关于Y的陈述。这个X与Y之间的映射、这一类比、这种部分的同构性——你想怎么叫它都无所谓——高效而可靠地把意义从一个框架带入了另一个框架之中。

让我们再到一个更加微妙的浪漫情境里去看看这种交流的模式。奥德丽不太确定本对自己认真到什么程度，于是“不经意地”把聊天的话题转移到他们共同的朋友辛西娅和戴夫的关系上，又“不经意地”问本，他对于戴夫无法对辛西娅做出承诺这件事怎么看。本也不傻，他迅速地察觉到了此中的危险，所以在一开始的时候，他小心翼翼，不敢给出任何具体的见解，担心在“单纯”谈论戴夫时把自己也牵连进去，但是他随后又意识到，这份危险也为他提供了一个机会，可以向辛西娅传达一些他一直以来都不敢直接提起的事情。打定主意后，本便以一种精心装扮出来的漠然语调回应说，他可以想象戴夫为什么在做出承诺这件事上迟疑不决，因为说到底，辛西娅还是比戴夫的才智高出太多了。本希望奥德丽可以明白其中的暗示：既然她也在对艺术的投入上比他深沉太多，所以这便也是他在对她做出承诺这件事上同样迟疑不决的原因。借助伴侣与伴侣之间相当强烈的类比，他的暗示含蓄而又清晰地传给了她。这个类比已经在本和奥德丽的头脑中酝酿了几个月的时间，只不过未曾传达。由此，本设法在没有直接谈论自己的条件下，清楚地表明了自己的态度，而且他和奥德丽两个人都心知肚明。

上面这个场景也许会让你觉得太做作了，因而给你带来的印象大概是：把一段恋爱关系视为另一段恋爱关系的“编码”，是一种脆弱不堪且不太现实的可能。但是这并不意味着远离了真相。如果两个人处于恋爱关系之中（或者哪怕他们并非情侣，只要其中一个人感觉到有潜在的火花即可），那么他们两个人之间有关任何一段浪漫关系的任何一场谈话，不论关于谁，都很有可能被两个人中的一个或两个理解成是对他们

自己关系的一个投影。这种回转反噬几乎是无可避免的，因为浪漫关系总是充满了渴望与不确定性，即便是非常好的浪漫关系也不例外。我们总是在密切留意着投向我们恋爱生活的线索与洞见，而类比则为线索与洞见提供了最丰沛的资源。因此，注意到我们自己与另一对处于话题中心的伴侣之间的类比，对我们而言简直是小菜一碟。关键的问题在于，这碟菜的味道好不好。

乡村面包师评语中潜在的歧义

刚刚讨论过的那一类间接指涉，常常在文学中被艺术化地加以利用。在文学作品里，如果读者能轻松地观察到情境A与情境B之间的一个强大的类比，他们也就能把情境A中的人物语言套用在情境B之上。有时候，情境A中的人物对于情境B中发生的事情一无所知，这就能造成一种幽默的效果，而另外一些时候，情境A中的人物也在情境B中出现，只不过他们不知道（或者没有想到）将他们所处的两个情境联系起来的类比。显然，后一种情况创造出了一种强烈的反讽。

我最近刚好遇到一则非常有趣的例子，忍不住想在这里讲给你听。它出现在马塞尔·帕尼奥尔（Marcel Pagnol）于1938年拍摄的电影《面包师的老婆》（*La Femme du boulanger*）的片末。当老婆奥蕾丽（Aurélie）与一个牧羊人私奔三天后灰溜溜地负疚归家时，名字滑稽的乡村面包师可亲（Amiable）迎接她的方式完全是宽宥而愉悦的——可是，他的猫，小美，恰好<u>也在</u>三天前离家，抛弃了她的配偶绒球，并且在奥蕾丽回家的同一天<u>也</u>回到了家中（当然，这些都纯属偶然），而可亲对她的态度却是相当的残酷无情。可亲与受伤的绒球同仇敌忾（有些人会说他是“感同身受”），把小美痛骂到体无完肤，而这一切都发生在刚刚回到家的奥蕾丽眼前。他所使用的那些残酷的字句，都是观众们本以为是留给奥蕾丽用的。仿佛这还不够，奥蕾丽吃掉了可亲为自己晚饭准备的心形面包（他根本不知道她会回来），而迷途的小猫小美，脖子

上挂着一个镶有巨大心形的项圈，也吃掉了刚刚为她的配偶绒球备好的食物。

面包师可亲是否确实观察到了这个呼之欲出的类比呢？还是说，他的心地如此善良，又宽宏大量，看不出奥蕾丽和小美其实是一个豆荚里的两粒豌豆？他凶狠（而正当）地把怒火泼在小猫身上，这在我们听起来是一次酣畅的双向指责，而在他看来，这可能纯粹是单向的吗？

不管真实情况究竟如何，我强烈建议你亲自去看一遍那部电影；那是一部针针见血的杰作。而如果机缘奇巧，在你观看这部电影时，坐在你身边陪你一起欣赏的爱人恰好也刚刚结束一段与第三者的“露水情缘”而重返巢中，那么大可想象，当那最后一幕来临时，他或她将有多么如坐针毡！但是，为什么一个身处电影<u>之外</u>的人会因为由电影<u>之内</u>的人发表的一阵扑面而来的严厉谴责而感到扎心呢？这个嘛……类比的力量总是与它的精确度与可见性成正比的。

尚塔尔与意义的叠加层面

我们下面要讨论的这个类比，其双方比两块曲奇或一对恋人之间的差异更大，甚至比一个迷途的妻子和一只走失的母猫之间的差异还大。这个类比隐含地发生在我们在电视上观看一段视频的时候——比如说，一段关于一位法国面包师、他的老婆、他的朋友和他的猫的节目。重点在于，我们并不是<u>真的</u>在观看人和猫的玩耍嬉戏——至少严格说来不是。这是一种类比，只不过是一种实用的简便说法，因为我们真正在看的本来是大量的像素点，它们以一种完美的同步方式，动态地转换着色斑的模式，而这些散布的色斑原本展现的是很久以前一个遥远的法国村庄里的生命体和非生命体。我们正在观看的是上百万个“编码”了那些人行动的小点，但幸运的是，这种编码很容易就能被我们破解——完全不费吹灰之力。事实上，我们被这种映射或这种同构性（如果你愿意的

话，也可以说这是屏幕与场景之间的类比）深深地迷住了，感觉像是自己被“时空传送”到了某个遥远的地点与时间，似乎正在看到的就是事件本身发生的样子；我们觉得，严格区分我们是否“真的”在观看那些事件，简直是令人恼火的鸡蛋里挑骨头。（如果我们打电话聊天，那么我们真的是在同彼此讲话吗？）

人们很容易忘记，飞蛾、苍蝇、狗、猫、新生儿、电视摄像头以及其他小灵魂的生命，是不会像我们这样观看电视屏幕的。虽然不太好想象，但是它们看到的是未经加工与阐释的像素，因而在它们看来，电视屏幕里并没有什么很久以前的遥远的意义，就如同在你我眼中的一堆落叶、杰克逊·波洛克（Jackson Pollock）的一幅画作或者一张马尔加什语的报纸一样没有意义可言。（如果你会说马尔加什语的话，那么我很抱歉，请换成冰岛语——你别告诉我连冰岛语你都会说！）在意义表征的层面上“阅读”电视屏幕远远超出了这些生物的智力水平，甚至对于大多数人类而言，这也是到了两岁左右才会发展出来的后天本性。

一条狗茫然地盯着电视屏幕，无法辨识出任何图像，甚至都没有意识到它本来意在传达图像，这与罗素公爵的情况并无不同。他茫然地盯着自己深爱的*PM*系统中的一个公式，只看到了它的“简单”（算术）意义，而其他的意义，由哥德尔发现的经由映射而产生的意义，则超越了他的智力水平，完全不可触及，做梦也不曾想到或许你会觉得，这种比较对于伯特兰爵士来说太不厚道了，而我多少也表示同意，所以下面请让我用一种更加切实也更加宽容的方式重新加以说明。

暂且忘记那条在电视机屏幕上只能看到像素而看不见人的狗，让我们想象一个三岁的小女孩尚塔尔·杜普莱西思科，她正在和父母一起观看《面包师的老婆》。三位观众的母语都是法语，所以这里没有语言上的障碍。和她的妈妈和爸爸一样，尚塔尔也透过像素直接看到了村庄里的事件，而当那场绝妙的最后一幕来临，可亲开始大肆谴责那只猫的时候，尚塔尔大笑起来，而引她发笑的是可亲的怒火——但是她一刻也

没有怀疑过，他的话里还有另一层意思。她的年纪太小了，还不能理解奥蕾丽和小美之间的类比，所以对她而言，这里只有一种意义。电影作者帕尼奥尔经由类比而制造的意义，以像素与遥远事件之间的“简单”（虽然狗捕捉不到）映射为默认的前提，因而是在此之上叠加的一层。尚塔尔的父母可以毫不费力地观察到这层意义，但是就目前来说，它还处于尚塔尔的智力范畴之外，所以她是完全无法触及的。当然，几年之后就不一样了——尚塔尔将学会如何拾取各种各样复杂情境之间的类比。可是现在，事情就是这样。

我们可以拿这个情境与伯特兰·罗素进行一个更加切实和宽容的比较。（这又是另一个类比！）尚塔尔和那只狗不一样，她看到的不仅仅是屏幕上无意义的光的图案，而是可以毫不费力地看到人和事——这些图案的“简单”意义。但是还有把这些人和事当作天然前提的另一层意义，由事件之间的类比传递出来，而这正是尚塔尔捕捉不到的更高层面的意义。几乎是以同样的形式，伯特兰·罗素也捕捉不到哥德尔的更高层级的意义，经由他的映射，即他那非凡的类比，所调制的意义。就我对罗素的阅读所见，他从未看到过*PM*公式的这第二层意义。令人唏嘘的是，这位伟大的公爵从来都没有学会阅读由他自己创造的圣书。

时装店的罢工警戒线

我在前文提到过，你那在外浪荡的心肝宝贝刚刚回家时，大抵该听得出可亲对小美的责备里隐含的额外层面的意义。因而，一部戏剧或电影可能承载着作者做梦也未曾想到过的意义。例如，让我们来看看一部在1931年创作的鲜为人知的戏剧，《时尚商店大罢工》（*The Posh Shop Picketeers*），作者是社会活动剧作家罗莎琳·瓦德海德（Rosalyn Wadhead），你听说过她吗？这部戏剧讲述的是由阿尔夫和伯蒂的时尚商店（我得承认，我到现在都没搞清楚他们在那里卖的是什么东西）里的工人们发动的一场突然的罢工运动。戏剧中有这样一幕，走向商店入

口的购物者们被劝说不要越过警戒线，也不要在商店里买任何东西。（“阿尔夫和伯蒂肮脏下流！请不要越过我们的时尚商店警戒线！请去马路对面的夫妻店！”）在我们剧作家精湛的技艺之下，这个简单的情境发展成一出充满张力的戏码。不成想，出于某种原因，就在这出戏剧即将上演之前，剧场的引座员和剧中的演员卷入了一场激烈的冲突之中，结果引座员联盟在首演当夜发动了一场突袭的罢工，竖起了警戒线，并恳求有心观戏的人不要跨过他们的警戒线去观看《时尚商店大罢工》。

显然，在这一意想不到的政治语境之下，剧中演员的台词为座席上的观众呈现了一种强大的第二种意义，这是罗莎琳·瓦德海德未曾料到的一种额外层面的意义。剧中有一位参与罢工的时尚商店店员名叫“凯吉”，在一位傲慢无礼的妇人把她一把推开、大步流星地迈入阿尔夫和伯蒂的高档陈列室之后，她满脸厌恶地宣称：“每一个越过阿尔夫和伯蒂时尚商店门前警戒线的人都是人渣。”事实上，观众席上的每个人都不可避免地听到了这句话（而这些人注定只包括那些跨过了剧场门前警戒线而进来看戏的人），而在他们的耳朵里，这句话无疑变成了：“每一个越过剧场门前警戒线的人都是人渣”，这当然也等于在说：“每一个人坐在观众席上的人都是人渣”，也可以等同于在说“你不应该听这些台词”。然而，这与台上所有的演员想要告诉台下观众的话截然相反，也包括饰演凯吉的那位演员。这些观众不顾引座员恶意的警戒线，走进剧场来观看这场戏剧，演员们自是对此由衷地感激。

但是，这些演员却明白无误地把他们深深感激的观众叫成了“人渣”，还间接地暗示出，甚至根本不该有任何人听到这些台词。面对这一事实，演员们能有什么办法呢？只能束手无策。他们*不得不*背诵出剧本上的台词，而类比如影随形，直白而强烈。因此，凯吉的那句台词，以及其他很多句台词，其中蕴含的反讽、扭转、自我指涉的意义，都无可避免。我们可以承认，这种自我指涉是*间接的*——以类比为中介，但是它的真实和强大丝毫没有因此而逊色于“直接”的指涉。实际上，那些

更有可能被我们称为“直接”的指涉同样也需要一套编码作为中介——由我们的母语（马尔加什语、冰岛语等等）提供给我们的语言与事物之间的编码。只不过，那种编码是一种更加简单的编码而已（或者至少是更加熟悉的一种）。总而言之，“直接”与“间接”的指涉之间看似截然分明的区别，也不过是程度上的差异，而不是非黑即白的区分。再重申一遍，类比的力量与它的精确度和可见性成正比。

原理王子：数学戏剧

好吧，关于罗莎琳·瓦德海德和《时尚商店大罢工》中那句意料之外的双刃剑台词，就说到这里为止。不得不承认，这部戏剧实在是太默默无闻了。让我们来点儿完全不一样的东西吧。我们下面要讨论的是一出世界闻名的戏剧，由大名鼎鼎的英国剧作家（你总该听说过他吧！）怀·泰德·因罗素（Y. Ted Enrustle）于1910年到1913年间创作的《原理王子：数学戏剧》（*Prince Hyppia: Math Dramatica*）。这位剧作家受够了当时盛行的那种“抖机灵抖过了头”的关于戏剧的戏剧，于是起笔写成一部与剧本创作、表演或舞台不会产生任何牵连的戏剧。因此，在部闻名遐迩的著作中，所有的人物角色都受到了严格的限制，你肯定还记得，他们可以谈论的只有自然数的各种不同特性，从最简单的到最深奥的特性都可以，但是仅限于此。还有比这更不像“关于戏剧的戏剧”的创作吗？让我们看看具体的例子。在第一幕刚开始不久，美丽的表述公主发表了一句著名的宣言：“7乘以11乘以13等于1 001！”而帅气的原理王子则激动地回嘴道：“正因如此，1 001是个合数，而不是一个质数！”他们口中的数学仿佛来自天堂一般（现在，你可能要发出咆哮了）。

但是，事情真的发展到白热化还是在第三幕。当表述公主提到了有关某个超大整数*g*在算术上的事实时，高潮来临，原理王子给出了回应：“正因如此，*g*这个数字是粗俗数，而不是质雅数！”（在听到原理王子这最数学戏剧化的爆发后，极少有观众不会跟着大家一起倒抽一口

凉气。）奇妙的是，骄傲的原理王子似乎对他口中台词的重要性一无所知，甚至更具反讽意味的是，剧作者怀·泰德·因罗素显然也不知道。不管怎么说，时至今日，我们每个人都知道，原理王子的评论——经由一个紧密的类比中介链——声明的是，基于那个已逝时代所秉持的标准戏剧规范，某个长长的排印台词是“不可言说”的。而有趣的是，那位扮演原理王子的演员刚刚发声宣告的，正是他自己所称的不可言说之语！

你大可想象出，虽然怀·泰德·因罗素在书写长长的台词时，向来谨守着通行的戏剧规范（毕竟他要靠这个养家糊口），但是他连做梦也未曾想到过，在自然数（他笔下的神奇角色可以精准地说出这些数字的怪异特征）和他写给演员去阅读和背诵的平实的台词之间还存在着一种联系。将近过了20年之后，一位年轻而尖刻的土耳其裔维也纳剧评家戈尔德·库洛德写了一篇题目为“论《原理王子：数学戏剧》中在形式上不可言说的命题及相关舞台剧（一）”的活泼锋利的剧评，才把这一古怪的巧合公之于众（具体的细节在此跳过不提，因为这个故事已经人尽皆知了）。这篇剧评锋利而中肯，立即得到了众人的赞赏，由此带来的结果是，读过库洛德这篇不敬的评论的观剧者，都可以从《原理王子：数学戏剧》中很多著名台词中重新听出怀·泰德·因罗素本意之外的意义，仿佛这些台词讨论的根本不是数字，而是对怀·泰德·因罗素戏剧本身的直接（而且常常是相当尖锐的）评论！

因此，不久之后，懂行的观众便开始对不合群的数字命理学家鬼多计（《原理王子：数学戏剧》中的一个人物，因为总是喋喋不休地唠叨她为什么更喜欢粗俗数而不喜欢质雅数而臭名昭著。）的古怪评论展开了重新的解读，并发现她的评论经由如今看来再明白不过的影射，其实揭示的是她为什么更喜欢（使用当时的戏剧规范）不可言说的戏剧台词，而不喜欢可以付诸笔端的台词。对于戏剧爱好者而言，理解这部戏剧的新方式简直妙不可言，因为它（带着复仇的快意！）揭示出《原理王子》其实是一部关于戏剧的戏剧。但是不要忘了，得出这份洞见的绝大部分功劳都属于那位鲁莽的年轻外国评论家，而跟那位饱受尊敬的前

辈剧作家没有多大关系。

怀·泰德·因罗素，可怜的家伙，此时只能目瞪口呆——也没有别的词可以形容他了。怎么可能有任何理智正常的人，会用这种荒诞不经的方式来解读鬼多计的台词呢？它们只是在讨论数字！说到底，写成一部有关且仅仅有关数字的戏剧，从来都是他唯一的宏图，而他为了完成那个高贵的目标已经勤勤恳恳地耕耘了多年！

怀·泰德·因罗素以文字发表了猛烈的还击，坚称他的作品绝不是一部关于戏剧的戏剧，更别说是谈论自身的戏剧了！事实上，他还不惜更进一步坚持认为戈尔德·库洛德的评论对象怎么想都不可能是《原理王子：数学戏剧》，他评论的肯定是另一部戏剧，可能是一部相关的戏剧，也许是一个类似的戏剧，甚或是一部完全类似的戏剧，也有可能是由两个偏执的悖论恐惧症患者所创作的一部题目相似的戏剧，但无论如何，都不可能是他的戏剧。

然而，不管如何抗议，面对如今的观众解读他挚爱的戏剧台词的方式，怀·泰德·因罗素都无力回天，因为受到启发的观剧者已经完全把那两个概念——某些整数的粗俗数和某些戏剧对话台词的不可言说性——视为了精确同构的现象（与奥蕾丽和小美的离家出走之间的同构性毫无不同）。调皮顽劣的库洛德发现了这种微妙的映射，并欣然在自己的剧评中将其公开出来，（至少对那些读懂了评论的人而言）这就让两种意义得到了同样清晰的呈现。这里最大的反讽之处在于，对于一些算术评论，比如原理王子那次有名的爆发，将其所言之物听成是戏剧中不可言说的台词，比将其所指之物听成是非质雅数，要更容易，也更自然！可怀·泰德·因罗素尽管反复读过好几遍库洛德的评论，显然从未跟上真正的思路。

类比：做凯吉做过的事

好了好了，适可而止吧。玩笑开完了！让我坦白吧。在前面这几页里，我一直在玩一个游戏，谈论一个名字奇怪的剧作家创作的一部题目奇怪的戏剧，以及一个名字奇怪的剧评家写的一篇标题奇怪的评论，但真相（亲爱的读者，我想你早就知道了）是，我真正在讨论的其实是完全不同的对象——明确地说，就是奥地利逻辑学家库尔特·哥德尔（戈尔德·库洛德）在罗素和怀特海的《数学原理》中发现并揭示出来的那个怪圈。

“得了吧，”我听到有反对的声音（但是当然这不是你的声音）说，“如果你写下的那些文字都是关于怀·泰德·因罗素和他的《原理王子：数学戏剧》的，你真正谈论的东西怎么可能是怀特海和罗素的《数学原理》呢？”好吧，再一次，这又是得益于类比的力量；它和影射小说玩的是同一种游戏。小说家表面上宣称是在写纯粹虚构的人物，可真正在讲的却是现实生活中的人，而且不怎么多加掩饰，于是头脑清醒的读者可以精确地猜出谁是在代表谁，由于小说中的类比是如此强势而抢眼，在特定的文化语境中，但凡有足够阅历的人，都不可能对此视而不见。

由此，我们已经沿着我的梯子，从在咖啡馆脱口而出、转身即弃的“真难吃”一路向上，爬到了具有两种说法的评论的例子处，遇见了超级精密复杂的戏剧台词“数字g不是质雅数”。我们一次又一次地看到类比与映射是如何在初级意义上引申出第二种意义。我们看到，即便是初级的意义，也同样依赖于不言自明的映射，所以说到底，我们看到的是，所有的意义都是以映射为中介的，这也就是说，所有的意义都来自类比。哥德尔的这一深刻洞见，在他写于1931年的文章中发挥到了极致，把深埋在《数学原理》中的启示从地下翻出，令其得见天日。至此，我希望我所有的读者都可以不费吹灰之力地理解哥德尔对于意义的敏锐洞见了。

如何才能写出一句“不可言说”的台词？

当你得知原理王子那句关于数字g的著名台词（经由类比）声明了它本身的不可言说性时，不免会产生某些困扰。这难道不是自相矛盾的吗？如果某个剧本中的某句台词真的是不可言说的，那么剧作家怎么可能把它写出来呢？或者，把问题反过来问，如果原理王子的经典台词根本就没有被写出来过，那么我们怎么可能在怀·泰德·因罗素的戏剧中遇到它呢？

这的确是个好问题。但是别忙，请先回忆一下，我把“可以言说的台词”定义为一句可以由默默恪守一套现成的戏剧规范的剧作家写下来的台词。换句话说，“可言说性”内在地指向某种特别的规则系统。这就意味着一句“不可言说”的台词，并不是一句永远绝不可能由任何人写出来的台词，而只不过是违反了一条或几条大多数剧作家公认的戏剧写作规范的台词。因此，一句不可言说的台词其实是可以写出来的——只是要由某个不那么严格遵守规则的人来写。

严格受制于规则约束的剧作家写出这样一句台词，将被视为极端的不可理喻；如果遇到一位总爱变换花样恶意中伤的小肚鸡肠的剧评人，甚至可能得到这样的评语：“某人的戏剧实在是狗屁不通！”[注]也许正是因为察觉到了怀·泰德·因罗素的戏剧中出乎意料而又奇怪万分的“狗屁不通”之处，观众们才会在听到原理王子那番数学戏剧化的爆发时倒吸一口凉气。难怪戈尔德·库洛特会因为指出一句原先不可言说的台词已被写出，而赢得满堂喝彩。

“不”不是怪圈的来源

可能会有读者总结出，一个怪圈必然包括一种自挖墙脚或者自我否定的品质（“这个公式是不可证的”、“这句台词是不可言说的”、“你不应该观看这场戏”）。然而，否定在怪圈中并没有扮演任何至关重要的角色。只不过，如果怪圈享有一种自挖墙脚的品质，它的怪就会变得更

加刺激或好玩。回头想想艾舍尔的《画手》。里面不存在任何否定——两只手都在画。可以想象一下，如果一只手在擦掉另一只手，会发生什么！

在本书中，一个环路的怪异单纯地来自一个系统“吞食自我”的方式，即通过一种意料之外的回路扭转，粗暴地违反被我们认定为不可侵犯的等级秩序。不管是在《原理王子：数学戏剧》还是《数学原理》中，我们都看到了，一个精心设计成只谈论数字而不谈论自己的系统，最终还是难逃以“凯吉”的方式谈论自己的命运——而这恰恰是缘于数字反复无常的本性。数字的形式太丰富、太复杂，数字模式的灵活性令其可以镜照任何一种其他类型的模式。

如果哥德尔当初构造的是一个自我肯定的公式，并令其狂妄自大地宣称，“本公式在*PM*规则下是可证的”，那么他依然可以创造出一个同样怪的怪圈。虽然它也许会在戏剧性上略输一筹，但还是能让我想起不可一世的“我是最伟大的”穆罕默德·阿里（Muhammad Ali）以及“伟大的”萨尔瓦多·达利（Salvador Dali）。在哥德尔的身后，也的确[注]涌现出马丁·雨果·略布（Martin Hugo Löb）和里昂·亨金（Leon Henkin）等逻辑学家，构造并研究了这类自我肯定的公式。这些公式同样具有奇妙而深刻的特性。因此，我要再次重申，怪圈并非栖居于由“不”这个词导致的翻转之中，而是取决于包含了“本”这个词的违反了等级秩序的意料之外的向后扭转。

然而，我应该立即指出的是，像“本公式”这样的短语，在哥德尔的凯吉式公式中是找不到的——正如“本观众”这个短语也没有包含在“每一个越过阿尔夫和伯蒂时尚商店门前警戒线的人都是人渣”这句凯吉的台词之中。“现场的观众都是人渣”这层意想不到的含义，其实是由一个张扬显眼的类比所带来的不可避免的产物。这个类比即为发生在两条完全不同的警戒线（一条是在剧场外，一条是在舞台上）之间的映射，并由此引申为跨越警戒线的观众与在他们正在观看的戏剧中跨越警戒线的

越界者之间的类比。

认为像“本”（或者“我”“此处”“现在”——哲学家们把它们称为“索引词”——明确地指向说话者或者与说话者或信息本身紧密联系的东西）这样容易引发嫌疑的词是在一个系统中引发自我指涉的不可或缺的元素，这是一种先入之见，而哥德尔的发现已经表明，这只是天真的错觉而已；事实恰恰相反，这种奇怪的扭转只是由两种不同情境（一方面是被谈论的对象，另一方面则谈论者）的同构性所导致的简单而自然的结果。伯特兰·罗素确保了一切像“本”这样的索引概念都被杜绝在他的形式系统之外，因而深信自己的手艺已经可以一劳永逸地豁免于自我缠结的灾难——可库尔特·哥德尔却凭借其杀人诛心的同构主义表明，这种信念只不过是一种不合理的信仰。

作为一种表征中介的数字

为什么只有当某个人仔细钻研《数学原理》之时，这种同构主义才首次跃出水面呢？为什么在哥德尔之前，没有任何人想到过这件事？它之所以跃出水面，是因为《数学原理》在本质上是关于自然数的，而哥德尔之所见乃是，自然数的世界是如此丰富，以至于随便给出一类对象的某种模式，都可以找到与其同构的一组数字——换句话说，存在可以完美镜照一对象及其模式的数字，而这些数字将严格地踩着该模式的舞步起舞。而同步起舞才是此中的关键。

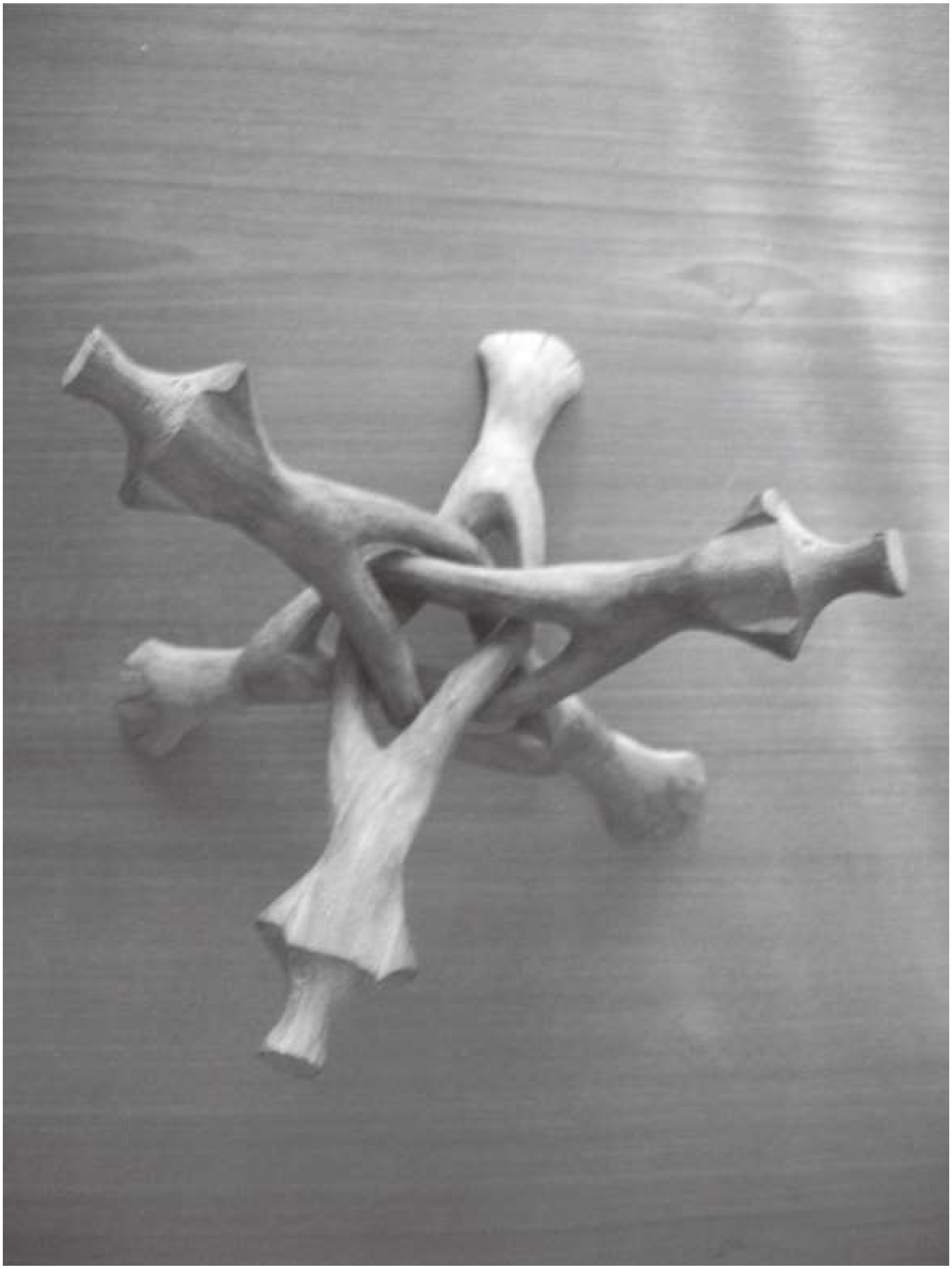
尽管从表面上看，正整数似乎相当朴素，而且彼此孤立，互不相关，可它们实际上却构成了一套极为丰富的表征中介，而库尔特·哥德尔则是首位意识到并利用了这一事实之人。正整数可以模仿或镜照所有类型的模式。就像在任何一种人类语言中，名词和动词（以及种种词类）都可以组织出无限复杂的舞步，自然数也可以编织出无限复杂的、累次相加相乘的舞步，因而也可以经由编码或类比而“开口”谈论任意种

类的事件，而不限于数字内部。正因如此，我才会在第9章中写道，自然数不论具有多么微妙的特性，*PM*都拥有足够的力量去谈论它，而正是这一看似无害的事实，蕴示了*PM*毁灭的种子。

生活在更早时代的人们，已经在很大程度上凭直觉感受到了这种丰富性：他们把我们周遭世界的各种不同方面——恒星、行星、原子、分子、颜色、曲线、音符、和弦、旋律以及种种——的自然本性，努力嵌入到数学方程或者其他数字模式当中。四个世纪之前，这一整个倾向最初的开创者伽利略发表了著名的宣言："自然之书是以数学的语言写成的"（这个想法在热爱自然而讨厌数学的人听起来，必定是触目惊心）。在随后的几个世纪里，人们对于世界不同方面的数学化获得了伟大的成功，可尽管如此，在哥德尔之前还从未有人意识到，数学本身的行为也是数学可以模仿的范畴之一。

由此可见，此中最基本的要点在于，潜行在《数学原理》内部而被哥德尔发现的那种意料之外的自我指涉扭转，乃是由自然数深藏不露的表征力量所造成的不可避免的自然后果。一个视频系统可以创造出一个自我指涉的环路，这并非什么奇迹，而显然是由电视摄像头的力量（准确地说，是巨大的像素矩阵所具有的极端丰富的表征力量）带来的一种零碎副产物；同理，《数学原理》（或者任何其他可与其相比的系统）中包括了如哥德尔公式一样关注自身的句子，也不是什么奇迹，因为整数系统与一个电视摄像头都可以"指向"任何系统，并能把那个系统的模式完美地复现在由它的一系列定理构成的那台隐喻的"屏幕"之上。而且正如在视频反馈中一样，来自*PM*的指向自身的旋涡也具有多种多样意想不到的呈展特性。想对其加以描述，需要一套全新的词汇才行。

1. 哥德尔在他的文章标题末尾放了一个罗马数字，这是因为他担心他没有足够明确地



表达出自己的观点，并认为他不得不再做一篇续篇。令他始料未及的是，他的文章很快便受到了来自约翰·冯·诺依曼和其他几位由名望的前辈的高度赞许。这让原本默默无名的哥德尔在很短的时间内声名鹊起。尽管如此，等到大多数的数学界人士都能理解他的成果的意义时，也已经过了几十年。

2. See [Hofstadter 2001] and [Sander], as well as Chapter 24 in [Hofstadter 1985] and [Hofstadter and FARG].

3. This should be heard as “X’s play is omegainconsistent”, which makes a phonetic hat-tip to the metamathematical concepts of *omega-inconsistency* and *omega-incompleteness*, discussed in many books in the Bibliography, such as [DeLong], [Nagel and Newman], [Hofstadter 1979], [Smullyan 1992], [Boolos and Jeffrey], and others. For our more modest purposes here, however, it suffices to know that this “o”-containing quip, plus the one two lines below it, is a play on words.

4. See [Smullyan 1992], [Boolos and Jeffrey], and [Wolf].

第12章
论向下的因果性

伯特兰·罗素的终极噩梦

在我的心目中，库尔特·哥德尔1931年作品中最出人意料的呈展现象，就是一种奇怪的数学因果性（假设我可以使用这个不寻常的术语）的新类型。我还从未见过有其他的评论者从这个角度来看待他的发现，所以下面所言纯属我个人的解读。为了说明我的观点，我不得不重新说回哥德尔那个著名的公式——为了纪念他，让我们把这个公式命名为“酷哥”（KG），并分析它的存在之于*PM*的意义。

正如我们在第10章的末尾所见，如果把酷哥的意义（或者更加准确地说，是其第二层意义——由哥德尔的机巧映射所揭示出来的更高层级、非数字且非罗素主义的意义）浓缩成一句精华，便是那句鞭子一样的陈述：“酷哥在*PM*内部是不可证的。”随之而来的一个自然的问题——那个自然的问题——便是：“好吧，那么，酷哥在*PM*内部真的是不可证的吗？”

要想回答这个问题，我们不得不首先笃定一种信念，即*PM*内部任何可证的东西都是真的陈述（或者倒过来说，在*PM*中，没有任何假的东西是可证的）。这种令人愉快的情形，就是我们在第10章中所说的“一致性”。如果*PM*是不一致的，那么它就会证明大量有关自然数的虚假命题，因为一旦你证明了任何一个特定的假命题（比如，“0=1”），那么无数的其他假命题（“1=2”，“0=2”，“1+1=1”，“1+1=3”，“2+2=5”，等等）也会遵循着*PM*

的规则接踵而至。事实上，情况远比这更糟：不管多么晦涩和难解，只要任意一种假陈述在*PM*中是可证的，那么每一种可以构想出的算术陈述，不论其真假，都会变成可证的陈述，而这座宏伟的大厦便会在令人惋惜的混乱中彻底崩塌。简言之，即便只有一个假命题是可证的，也可凭此判定*PM*与算术上的真理毫不相干。

那么，伯特兰·罗素的终极噩梦到底是什么呢？那便是：某一天、某个人、为某个公式做出了*PM*的证明，可那个公式却表达了某个不真实的算术陈述（“0=s0”是个不错的例子）。这件事一旦发生，*PM*便会立即无路可走，只能投奔到垃圾箱里去了。然而，罗素是幸运的，因为地球上每一位逻辑学家都愿意打赌，一个雪球在火狱里滚动百年而不融化的可能性，都比这件事发生的概率大。换句话说，伯特兰·罗素的终极噩梦终归仅仅是一个噩梦而已，绝不会成真。

为什么逻辑学家和数学家——不仅仅是罗素，而是（包括哥德尔在内的）所有人——都愿意下此赌注呢？(注)首先，*PM*中的公理肯定是真的，而且它的推理规则又是如此简明而合理，没有人能想象出比其更加稳妥坚实的状态了。你怎么可能从中得出假的结论呢？毫不夸张地说，想象*PM*可能容纳虚假的定理，跟想象2加2等于5一样困难。因此，让我们同所有数学家和逻辑学家一道，姑且给出对罗素和怀特海有利的裁断——假定他们宏伟的逻辑大厦具有一致性。于是，自此以后，我们都将大方地假定，*PM*绝不会证明任何假的陈述，它所有的定理都必然是真的陈述。然后，在我们这一友善假定的前提下，让我们问问自己：“如果酷哥在*PM*内部是可证的，那么接下来将有何事发生？”

“因为”与“虽然”偶遇的奇异大陆

就是现在，读者，让我们设定酷哥在*PM*中是可证的，然后看看这个假设——我将其命名为“酷哥可证设想”——会把我们引向何处。请注

意，这里的反讽之处在于，酷哥本人并不相信酷哥可证设想。他倔强任性地向全世界喊道："我不是可证的！"所以，亲爱的读者，如果我们关于酷哥的看法是正确的，那么他关于自己的说法就是错误的，不管他的呐喊有多么响亮。毕竟，一个公式不可能同时（如我们所称）可证并（如酷哥所称）不可证。我们其中之一必定是错的。（而且对于任何一个公式来说，错就意味着假。）所以……如果酷哥可证设想是成立的，那么酷哥就是错的（假的）。

就是这样。我们的推理从酷哥可证设想出发，最后得到的结论是"酷哥是假的"。换句话说，如果酷哥是可证的，那么它同时也是假的。但是，且慢，这岂不是——一个在*PM*中可证的虚假命题？！就在片刻之前，我们不是斩钉截铁地声明，*PM*绝不会证明假命题吗？没错，我们确实做出了声明。我们赞同逻辑学家普遍的观点，认定*PM*是具有一致性的。如果我们坚守己见，那么酷哥可证设想便只能是错的，因为它所引向的是罗素的终极噩梦。我们不得不收回它、取消它、否认它、废除它并撤销它，因为如果接受它，我们便会被引向一个我们已知是错误的结论（*PM*不具有一致性）。

由此一来，酷哥可证设想就在此被推翻了，而留给我们的是与其相反的一个设想：酷哥不是可证的。滑稽的是，这恰恰是酷哥朝着天花板大喊的话。我们看到，酷哥对自身的声明——"我是不可证的！"——是真的。简而言之，我们认定了两个事实：（1）酷哥在*PM*中不可证；（2）酷哥是真的。

我们刚刚发现的是*PM*内部一个十分奇怪的地方：这是一个我们可以确定其为真的算术（或者说得更精确一点，是数论）陈述，然而我们还同等地确信它是不可证的——雪上加霜的是，这两个听起来相互矛盾的事实还是相互导致的结果！换言之，不仅可以说：酷哥虽然是真的，但是不可证；而且更糟的是，酷哥正因为是真的，所以才不可证。注

这种奇怪的情况完全是史无前例的，而且相当悖于常理。它当面挑衅了数学家的信条；后者主张，真实与可证性乃是同一枚硬币的两面——它们总是如影随形，因为它们离不开彼此。而我们刚刚遭遇的情况恰恰与此相反，令人吃惊的是，真实竟然与<u>不</u>可证性同流合污。这真是太尴尬了！

从强项中衍生出的不完全性

在*PM*中存在着一个不可证的数论真理。如果你还记得第9章的内容，就会知道，这一事实意味着*PM*是<u>不完全的</u>。它的内部有漏洞。（到目前为止，我们只看到一个漏洞——酷哥，但是还有更多。事实上，它们的个数是无限的）。有些<u>本该</u>可证的数论陈述并没有收入到*PM*巨大的证明之网中——它们从网眼中漏了出去。显然，这又是另一种噩梦——也许不像罗素的终极噩梦那样具有毁灭性，但是在某种意义上甚至更加阴险，令人心神不安。

1931年的逻辑学家和数学家们当然没有料到会出现这样的状况。没有一丝迹象表明，《数学原理》中公理和推理规则在任何方面有缺陷。恰恰相反，它们看上去几乎包含了任何人可能想到的有关数字的任何一种真实陈述。本书第10章曾引用过哥德尔写于1931年的那篇文章的开篇，哥德尔在那一段话中明确地表达了此意。不知道你还记不记得，他在说到《数学原理》和策梅洛-弗兰克尔的集合论时写道："这两个系统的覆盖范围极为广泛，我们今天在数学中使用的所有证明方法，都已经在系统中完成了形式化，也就是说，它们被简化成了少数的公理及相应的推理规则。"

哥德尔在此表明的论点几乎是当时的一个普遍认可的信条，所以，他在随后25页的文字中对于*PM*不完全性的揭示，真如晴空中划过一道霹雳，震破苍穹。

雪上加霜的是，哥德尔的结论并非取自*PM*中的弱点，而是源于它的强项。这个强项便是数字的灵活性，或者说“千变万化”（chameleonic），令它们的模式可以模拟出推理的模式。哥德尔利用了这一简单而又非凡的事实，即我们熟悉的自然数可以严格踩着我们不熟悉的*PM*符号模式的舞步而翩翩起舞。说得更确切一点，他发明的质雅数与可证的符号串难分彼此，而*PM*的天然强项之一便是，它有能力谈论质雅数。正因如此，它便也有能力谈论（编码的）自身。一言以蔽之，*PM*的表达力量造成了它的不完全性。这是多么美妙的一个反讽呀！

伯特兰·罗素的第二个噩梦

*PM*的任何丰富与改进（比如一个拥有更多公理或更多推理规则或二者同时增加的系统）都会与*PM*同等地展现出数字的灵活性（否则它就是更弱的，而不是更强的了），所以也会被同一种哥德尔式陷阱成功地捕获(注)——它同样也已经把自己绑在了自己的炸药箱之上。

让我们说得更具体一点。在这个更庞大的、据说更优越的超级*PM*系统中可证的符号串将会由一套比质雅数更丰富的数字集合（我们可将其称为“超级质雅数”）同构地模仿出来。然后，哥德尔立即可以按照他对待*PM*的同样方式，为超级*PM*创造一个新的公式“酷伙”（KH），其陈述的内容是“数字*h*不是一个超级质雅数”。当然，这里的h正是酷伙的哥德尔配数本身。（在*PM*中完成过这件事之后，再对超级*PM*做同样的事就轻而易举了。）我们刚才为*PM*逐步展开的推理模式将原封不动地再次重来一遍，而那个本应更强大的系统，也会以原封不动的方式，出于同样的原因，同*PM*一样，倒在不完全性的脚下。这正应了那句古老的谚语：“爬得越高，摔得越狠。”

换句话说，*PM*中（以及在任何一种跟*PM*同样丰富的公理系统中）

的漏洞并不是由罗素和怀特海的粗心大意造成的，而仅仅是一种任何灵活到足以涵盖自然数千变万化的品质的系统都摆脱不掉的特性。*PM*足够丰富，有能力像一个指向呈现着其所取画面的屏幕的电视摄像头一样，调转过来而指向自身。如果你的电视系统足够好，那么这种循环的能力就是不可避免的。而且，屏幕的分辨率越高，画面也就越真实。

这就像在柔道比赛里一样，你的对手的软肋正隐藏于他的力量之中；而哥德尔如一名黑带选手一样，使用*PM*自己的力量击倒了它。在不完全性上的击倒，并不如在不一致性上的击垮那样具有毁灭性，但这种击倒方式却完全出乎意料。任何扩大或丰富*PM*的方式都无法助你躲过哥德尔的黑带杀技，这一事实被称为“本质上的不完全性”（essential incompleteness）——伯特兰·罗素的第二个噩梦。但是与他的终极噩梦不同，第二个噩梦不再只是一个噩梦而已。它已经闯出了梦的领地。

络绎不绝的怪物

扩展*PM*于事无补；更糟的是，酷哥绝非*PM*中唯一的漏洞。任何一个给定的公理系统都有无限种哥德尔配数编码的方式，而每一种都能各自生产出酷哥的近亲。它们彼此不同，却又如此相像，简直如克隆体一般。如果你有心补天，大可以把酷哥或其任何一种克隆体作为公理丢入*PM*之中，（就此而言，你也完全可以把它们一股脑儿地全丢进去！）但是你的英雄壮举仍将无济于事；哥德尔的配方会立刻制造出一个全新的酷哥的近亲。然后，这个新的自我指涉的哥德尔符号串，还是与酷哥和他的诸多克隆体“一模一样”，而又与之前的每一个克隆体都不完全等同。你可以把这个公式也丢进去，可你又会得到另一个近亲！看起来，漏洞在顾此失彼的*PM*中层出不穷，如同按下葫芦浮起瓢。现在你该明白为什么我会说这个噩梦比罗素的终极噩梦更加阴险和令人心神不安了吧。

这一惊人反常却又极其美妙的招式，不仅打了罗素一个措手不及，而且几乎把每一位数学思想家都逼入了窘境，连伟大的德国数学家大卫·希尔伯特也不例外。要知道，他终其一生为之奋斗的主要目标之一，便是把所有的数学严格圈定在一个公理框架之内（这被称为“希尔伯特计划”[注]）。直到1931年这道晴天霹雳之前，人们都普遍地相信，怀特海和罗素已经实现了这一伟大的目标。

换一种方式说，那个时代的数学家普遍地信仰我在前文提过的“数学家信条”（《数学原理》版）。哥德尔石破天惊地揭示出，那看似合理的信仰的根基在根本上便具有不可弥补的缺陷。他是从以下两点推出这一结论的：一方面，我们友善地假定，这个基座是具有一致性的（即我们在*PM*的定理中绝不可能发现任何潜伏的假命题）；另一方面，考虑到*PM*的一致性，则酷哥及其无限数量的近亲在*PM*中都是不可证的，其原因如前文不远处所示，是根源于他们的自我指涉。

再重温最后一次，酷哥（或者他的任意一个近亲）到底是因为什么而不可证？一句话，就是因为它的自我指涉式意义：如果酷哥是可证的，它的循环性就会翻转过来，使其不可证，因此*PM*就是不一致的，而我们又知道它并非不可证。

但是，请注意，关于这些力图使酷哥作为底线来呈现的推导，我们还没有进行过任何详尽的分析。事实上，我们完全忽略了酷哥的罗素式含义（我一直将其称为它的原始意义），其声明了g这个庞大的数字具有一种相当晦涩而珍惜罕见的数论性质，即我所谓的“粗俗数性”或“非质雅数性”。你应该能注意到，前面的几页文字对于质雅数或非质雅数及其数论性质只字未提，也完全没有提及数字g。我们仅仅通过分析酷哥的第二意义，便机智地处理了所有这些与数字相关的问题；显然伯特兰·罗素从未真正理解这种意义。（本章第二节中）几行纯粹的非数字推理便足以说服我们相信，这个（关于数字的）陈述不可能令人信服地被认定为*PM*中的定理。

一致性宣判了峰顶的不可攀登

请想象，有一队乘坐卫星的探险员，刚刚发现了喜马拉雅山脉毫无疑问的最高峰（我们把它命名为“酷极”），并想象他们立即自信满满地宣告，由于这一峰顶具有某种特殊的极不寻常的特性，所以没有任何一条路线可以通到它的顶部。仅仅凭借从250英里之外的上空向下垂直拍摄的一张照片，这个小队就宣布了酷极是一个不可攀登的峰顶，而他们得出这样一个夸张的结论时，完全没有考虑从一个传统登山者的视角出发所见的峰顶的特性。更不用提亲身冒险、真正去尝试那数不清的可能路径中任何一条通往峰顶的陡坡了。“用不着，没有一条路是可通的！”他们欢欣地断言。“不需要去尝试任何一条路线——你肯定不会成功！”

如果这件古怪的事真的发生了，那么它便与之前得出的有关峰顶的可攀登的所有结论大相径庭。在此之前，登山者总是不得不尝试诸多路径——事实上，还要反复多次地尝试，在各种不同的天气条件下、使用各种不同的装备去尝试，哪怕连续上千次的失败也不足以构成某个峰顶不可攀登的铁证；人们唯一能够得出的结论是，它目前为止尚未被人类征服。没错，“不可攀登的证据”本身就是与登山这项活动最不相容的一个观念。

相反，我们的探险队根据酷极的某种全新特性得出结论，说它从本质上就是不可攀登的，而从未考虑过（更不用说亲身尝试过）诸多可能通向顶点的路线中的任何一种。尽管如此，据他们所称，他们的结论不仅是很可能或极有可能正确，而且是确凿无疑的。

这等同于提出了一个前所未有、冠履倒易、首尾错置的登山因果性。究竟是何种特性使这座特异的峰顶变得不可攀登呢？在每一条可能的路线上，登山者都将不可避免地遭遇某种不可逾越的致命阻碍，这样一种笼统的结论只能让传统的登山专家大惑不解。他们也许会相对谦虚

地总结说，如果从下往上看，并力图考虑所有可以想象到的登山路径，则攀登那个遥远的峰顶也许是极其困难的。但是我们这支无畏的团队正与此相反，我们看到的是酷极的顶端，并且自上而下地得出结论，就是根本不可能有任何一条路线可以通上来。

如果继续刨根问底、紧追不舍，这对探险员最终还是会说明他们到底是如何得出了这个粉碎了一切可能的结论的。原来，从上空拍摄的酷极的照片，所利用的并非平常的光线，而是一种新发现的“哥德尔光线”。当酷极经由这种全新媒介进入观察视野时，一套深藏的不可逾越的关键性结构便大白于天下了。

问题的根本在于支撑山顶冰川的岩石基座的稳定性；它太脆弱了，但凡有任何登山者闯入峰顶的受力范围，哪怕对其施加一点最轻微的重量（甚至于一颗盐粒；甚至于一只蜂鸟雏的睫毛！）都会立刻触发一场山崩地裂，而整座山都会倒塌陷落。所以，峰顶的不可攀登与一个人如何努力抵达毫无关系；而是关系到峰顶本身内在的不稳定性，不但如此，这还是一种只有哥德尔光线能够揭示出来的不稳定性。这真是一个傻乎乎的神奇幻想，不是吗？

数学的向下因果性

这确实有点傻。可是，库尔特·哥德尔的那颗重磅炸弹，虽然同样神奇，却并非只是一个幻想。它是严格而精确的，揭示出了一个令人咂舌的事实，即一个公式隐藏的意义也许拥有一种“向下”的因果力量来决定这个公式的真假（或者它在*PM*或任何一种足够丰富的公理系统内部的可推导性或不可推导性）。如此，则人们不需要再以故有的方式推导公式的真假，即按章守法地从公理出发，“自下而上”地艰苦跋涉——而是仅仅通过了解公式的意义，就可以推断出它的真假了。

这已经不仅是古怪了；它简直令人惊诧。正常情况下，一个人不可能仅仅看一看一个数学猜想所说的是什么，就可以单纯凭借那一陈述的内容本身而推断出这个陈述是真还是假（或者是可证的还是不可证的）。

例如，如果我告诉你，“完全数的个数是无限的”（完全数就是像6，28和496一样，都是因子相加之和等于其自身数值的数），你不会知道我的断言——可以叫它“小恶魔”——是真的还是假的。而且，仅仅是长时间地盯着写在纸面上的小恶魔的陈述（不管是用中文表达还是用某种如*PM*一样的烦琐的形式符号来表示），不会给你带来任何帮助。你将不得不尝试抵达这个峰顶的各种各样的路径。如此一来，你可能会发现，8 182是496之后的第二个完全数；你可能会注意到，你找到的完全数中没有奇数，这多少有点奇怪；你还有可能观察到，你发现的每一个完全数都可以写成$p(p+1)/2$的形式，其中p是一个奇质数（如3，7或31），而$p+1$同时还是2的幂数（如4，8或32）；你还会有更多类似的发现。

在经过一段时间以后，也许是在证明小恶魔的努力中经历了一系列的失败之后，你会渐渐地开始怀疑它是不是假的。在这种情况下，你也许会决定转变目标，开始尝试通往一个与其邻近的对峙高峰的不同路径，这一对峙峰顶即小恶魔的否定，～小恶魔，其所陈述的是“完全数的个数不是无限的”，也就相当于声称存在一个最大的完全数（这不免让人想起我们的老朋友p，据说其为世界上最大的质数）。

但是，设想你忽然之间灵感迸发，发现了一种新型的“哥德尔光线”（即某种机巧的新式哥德尔配数字编码，包括了令质雅数与可证串完美同步起舞的那种标准哥德尔装置的全部零件），令你可以看穿小恶魔，发现一个隐藏的第二层意义——这个隐藏的意义向有幸掌握了其解码方式的少数人宣称：“整数i不是质雅数”，而其中的i恰好是小恶魔的哥德尔配数本身。好了，亲爱的读者，我猜你不会认不出这个场景吧。

你应该很快就能意识到，小恶魔跟酷哥一样，通过你的新式哥德尔编码做出了有关自己的声明：“小恶魔在*PM*中不可证。”

在这种最令人欣喜却也最不可能的设想中(注)，你不再需要跑遍完全数以及其因数的世界或者跋涉在严格证明的世界中继续搜索了，你立即就可以得出结论说，小恶魔既是真的，也是不可证的。换句话说，你在得出“完全数的个数是无限的”这个陈述为真的结论的同时，还会得出另一个结论，认定无法使用*PM*的公理和推理规则证明它，而最终（转动反讽之刀）达成的结论是，小恶魔在*PM*中欠缺证明是由其真实性导致的直接后果。

也许你会认为我刚刚提出的设想是荒谬而不合情理的，可它就是哥德尔所做之事的一个精确类比。只不过，哥德尔没有从一个著名且有趣的关于数字的先在陈述入手，然后误打误撞地偶遇了其中隐藏的另外一种奇怪的意义，而是精心构造了一个关于数字的陈述，并通过他自己设计陈述的方式，揭示出其具有十分奇怪的另一种意义。但除了这一点以外，这两种设想是完全相同的。

假设性的小恶魔设想和真实的酷哥设想与数学的传统操作方式是根本不同的，而且我敢肯定你能说出它们之间的不同之处。它们等同于一种倒置的推理——从一个自命的定理自上而下推理，而不是由公理出发自下而上，特别是，它从一个自命定理的隐藏的意义出发，而不是由其表层的关于数字的声明开始。

格鲁大师，以及对真理机器的无功求索

你还记得格鲁大师吗？就是那台可以区分质雅数和粗俗数（非质雅数）的假想机器？我在第10章指出，如果我们造出了这样一台格鲁大师，或者如果干脆有人送给了我们一台，那么我们就可以判断出任何一

种数论猜想的真假。要想知道答案，我们只需要把猜想C翻译成一个*PM*公式，计算出它的哥德尔配数*c*（多么简单直白的工作），然后问格鲁大师：“*c*是质雅数还是粗俗数？”如果格鲁大师返回的答案为“*c*是质雅数”，那么我们就可以说：“既然*c*是质雅数，猜想C就是可证的，因而它就是真的。”而如果格鲁大师给出的答案为“*c*是粗俗数”，那么我们便可以说：“既然*c*是粗俗数，猜想C就是不可证的，因而它就是假的。”又因为格鲁大师（根据设定）总能最终给我们两个答案之一，所以我们大可以坐享其成，让它来解决我们能想象到的任何数学谜题，无论有多么深奥难解都没关系。

仅凭一个小小的装置便能解决所有的问题，这的确是一个伟大的设想；然而遗憾的是，我们如今已经可以看到，它有致命的缺陷。哥德尔为我们展示出，在*PM*中（事实上，在任何一种如*PM*一样的形式公理系统中），真实性与可证性之间隔着一道深远的鸿沟。也就是说，很多真的陈述是不可证的。所以，即便一个*PM*公式没能成为一个定理，你也无法将其视为它是一个假陈述的确定标志（可还算幸运的是，只要一个公式是一个定理，就可以确定它是真的）。所以即便格鲁大师名副其实，针对任何一个以“*n*是质雅数吗？”为形式的问题，总能给我们一个正确的“是”或“否”作为答案，它说到底也依旧无法为我们回答所有的数学问题。

尽管格鲁大师不如我们希望的那样强大，能够拥有一台这样的机器，依然是一件幸事，可真实情况是，连这份退而求其次的期望也不过是镜花水月而已。可靠的质雅数/粗俗数区分器，压根儿是不可能存在的。注（我不会在这里展开细节论证，但是你可以到很多数理逻辑或可计算性理论的教科书中找到答案。）突然之间，似乎所有的美梦都化成碎片，纷纷散落在我们的身边——而从某种意义上来说，这正是20世纪30年代发生之事。当时，真实性的抽象概念与确定真实性的机械手段之间的鸿沟首次得见天日，而这道鸿沟之大也渐渐开始为人知晓。

最后葬送这批数学家的这个梦想的人是逻辑学家阿尔弗雷德·塔斯基（Alfred Tarski）注。他指出，用*PM*的记号不可能表达出“*n*是一个真实成立的数论公式的哥德尔配数”这个文字陈述。塔斯基的发现意味着，虽然（使用哥德尔配数编码法）有无限多的数字可以用来代表真陈述，而且还有一套与其互补的个数同样无限的数字可以用来代表假陈述，但是想要用数论的形式表达出真假的区分，却绝无可能。换言之，所有的一整套*wff*都由真/假的二分法划分成两个互补的部分，但是真假之界线是如此怪异与无常，以至于绝不可能以任何数学的方式将其描述出来。

以上所有这一切看起来都太过怪异，可即便如此，这也是一种美轮美奂的怪异，因为它揭示了与人性同样悠久的数学目标之深邃奥妙。它展示出我们对于数学真理的共同求索其实是对某种微妙到难以言喻之物的求索，因而从某种意义上来说，这简直是对神圣之物的求索。我再次不由自主地想到，“哥德尔”（Gödel）这个名字的德语拼写中包含了“上帝”（God）的字样——而谁又知道，在字母ö上的两个点中还隐伏着多少不为人知的奥秘呢？

完成进化的生物，本末倒置的感知

我们在踏过的远征途中已经看到，数学逻辑中的怪圈具有十分惊人的特性，其中就包括一种以颠倒状态呈现的因果性注。但这绝不是我们在本书中第一次遭遇这种颠倒的因果性。这一概念在我们关于闹壳和人脑的讨论中早就冒出了芽。我们曾总结出，生物进化驯化人类去感知实体——把世界归为宏观范畴的实体。因此，我们生来注定不会以潜伏于表面之下的粒子物理学（这与我们日常的感知和熟悉的范畴相差了无数个量级）来描述身边发生的事，包括其他人所做之事以及我们自己所做之事。我们描述这些事时，使用的是抽象的、不甚明了的高层级模式，

如父亲和母亲、朋友和爱人、食品杂货店和结账柜台、肥皂剧和啤酒广告、怪人和天才、宗教和刻板印象、喜剧和悲剧、妄想与恐惧症，当然还有信仰与欲望、希望与恐惧、忧虑与梦想、野心与嫉妒、忠诚与憎恨，以及其他很多在隐喻的意义上与物理因果性的微观世界相距数百万英里的抽象模式。

因此，我们人类感知世界的正常方式具有一种奇妙的颠倒性：虽然微小的领域才是真正驱动现实的动因所在，我们却生来被打造成只会感知“大东西”而不会感知“小东西”。我们的思维完全无视低的层级而只能看见高的层级，这一事实不免令人想起哥德尔向我们呈现出来的那种高层级视野的可能性。他找到了一种处理绵长的*PM*公式（酷哥或其近亲）的方式，可以用一种简明的、易于理解的形式（“酷哥在*PM*中没有证明”）来解读它，而不需要再将其作为一种低层级的数字论断来进行解读，即某个特定的巨大整数具有某种特定的难解的递归式定义的数论特性（非质雅数性）。那种标准的对于一个*PM*串的低层级解读摆在明面之上，任何人都可以看到；而与之相反的是，只有天才出场，方能想象出还同时存在着一个高层级的解读方案。

对比而言，对于一个用大脑（或者一个闹壳）思考的生物，在一个更高的层级上解读它自己的大脑活动是相当自然的，完全不值得大惊小怪（比如，“我还记得那次祖母带我去看《绿野仙踪》时我有多么害怕”），而活跃在高层级之下的低层级活动（数不清的神经递质在突触间隙之间来回疯狂的跳跃，或者数十亿颗小磁交弹无声地相互撞击）却是完全隐晦的，既不会被想起，也不会被看见。一个思考的生物对于保障其思考得以发生的机制几乎一无所知，可尽管如此，它对于自己对世界的符号化解读却了然于心，而且还相当熟稔被其称为“我”的那个东西。

不论是好是坏，始终坚守在“我”的身边

实在罕有思考者愿意取消日常熟悉的符号并扔掉那个习以为常的“我”，进而大胆地推断出，在其大脑（或其闹壳）内部的某处可能存在某种难解的、隐匿的、更低的物理层级，充斥着某种肉眼看不见的翻滚和蒸腾，这与符号（或小磁交弹球）毫不相干，却以某种方式包含了数不胜数的微观单位，而且这些微观单位还以最神秘的方式欠缺一切属于符号的特质。

当你以这种方式思考人类的生命时，便会觉得我们早在从低层级的神经层面上意识到我们的大脑之前就已经在高层级、非物理的层面上（比如希望和信仰）意识到了大脑的存在则似乎是相当不可思议的一件事。（事实上，绝大多数人甚至从来没有在那个更低的层面上看待他们的大脑。）如果以《数学原理》作为类比，那么这便是说，对于某个特定*PM*公式的高层级的哥德尔式意义的认识，要远远发生在对其更基本的罗素式意义的认识之前，而这种情况看起来简直令人匪夷所思。虽然人类科学的整体发展在向后者迈出小小的碎步，可不管怎么说，我们人类就是进化出了以高层级的心理主义视角（“我希望明年夏天可以阅读《叶甫盖尼·奥涅金》”）的方式来感知和描述自我的天性，而不会采取低层级的物理主义角度（可以想象一下，促成你对于在第二年夏天阅读《叶甫盖尼·奥涅金》的希望的所有神经元状态，其数量将多到难以想象）。

向底层缓缓进军

当诸如“信念”“希望”“愧疚”“嫉妒”等心理概念业已发生数个世纪之后，方才有人想到可以尝试把它们解释为建立在某种物理基座（在某种极其微观的层面上观察到的活动的大脑）之上的反复出现的可识别的模式。这个趋势由尚且停留在高级层面上的直觉式理解出发，缓慢前进，逐渐发展到在低级层面上的科学化理解，这不免令人想起基因这一抽象概念推进的历史。最初只是有人大胆提出，父母把遗传特征传递给子

女，需要以某种基本单元为载体，而后又经历了数十年严谨的实验室研究，才有人发现了物理基础的“铁证”。当能够为这一抽象概念绘制一幅物理“图像”的微观结构终被发现时，结果却大出所料：基因原来是一条很长的螺旋扭线上中等长度的一段，只由四种分子（核苷酸）构成，它们依次相连，组成一条包括了数百万个单位的长链。

然后，如同奇迹一般，人们发现，从某种特定的意义上来说，这四种分子的化学性质原来并不重要——就有关遗传性的思考而言，最为要紧的是它们最新被揭示出来的信息特性，而不是它们传统的物理化学特性。也就是说，描述遗传和繁衍如何运作的恰当方式，在很大程度上可以从化学中抽象出来，仅仅保留一幅高层级的信息操作过程的图像[注]即可。

位于这些信息操作过程核心的，是一种高度抽象的“基因编码”概念，用每一种可能的核苷酸三联体“单词”，共计64种，来比对出同属于一个完全与此不相干的化学族（氨基酸）的20种不同分子。换句话说，要想深刻理解基因与遗传规律，必须将一种高层级的受意义左右的比对熟稔于心。这听起来应该并不陌生。

猪、狗与泥沼

如果你希望理解在一个生物细胞内部发生之事，就必须学会在这种新的信息层面上展开思考。虽然从理论上说，物理学是足够提供解释的，但是从可行性上说，单靠物理学来解释就是不行。很明显，那些基本的粒子悠然自得，根本不在乎生物分子的信息层面（更不用提人类的感知范畴、抽象信念、“我”、爱国热情或者燃烧的欲望了，也涉及不到一团特别的生物分子大结集或者创作一套24章的前奏与赋格）。宏观事件要从所有这些基本粒子的微观活动当中浮现出来，才能赐予生物以生命。

话虽如此，正如我在前文所指出的那样，如果你选择聚焦在粒子层面，那么你就没有办法划出一道清晰分明的界线，把如一个细胞或一头猪这样的实体从其所处世界中的其余部分中区分出来。像“细胞”或“猪”这样的概念在那么低的层面上是无关紧要的。粒子物理学的法则并不在乎如“猪“、“细胞”、“基因”或“基因编码”这样的概念，甚至对“氨基酸”的概念也不屑一顾。粒子物理学的法则只涉及粒子，而为了思考的便利性而画出的更大的宏观界线于它们而言是牛头不对马嘴的，就如同美国总统大选的分区界线之于蝴蝶的意义一样。电子、光子、中微子等等粒子，在穿越此类人工的界线时，根本不会感受到任何一点阻碍。

如果你走的是粒子路线，那么你就要贯穿始终，以这种方式处理整头猪，而遗憾的是，这同时也就意味着你将在远远的高度上飞越过那头猪。它要求考虑到这头猪的家族中所有成员、这头猪所居住的牲口棚、它在其中打滚的泥坑、喂养它的农夫、它呼吸的空气、落在其背上的雨滴、这些雨滴来源的积雨云、让这头猪的耳膜震动的雷电、整个地球、整个太阳以及弥漫在整个太空里并沿着时间延伸回宇宙大爆炸的宇宙背景射线中的所有粒子。对于我们这些有限的人类而言，这是一个大到无法完成的任务，所以我们不得不进行折中妥协，也就是在一个没有那么包罗万象也没有那么详尽的层面上来看待事物。（对我们而言相当幸运的是）这个层面反而是一个能给我们提供更多洞见的层面，而这便是信息的层面。

在这一层面上，生物学家可以讨论并思考基因代表了什么，而不仅仅聚焦在它们传统的物理化学特性之上。而且他们还潜在地接受了一个事实，即这种全新的、“减负”后的言说方式暗示出，基因得益于其信息特质，拥有了属于它们自己的因果特性——换句话说，特定的极端抽象的大规模事件或事态（例如，金毛猎犬性情温顺友好这一高层级的规律性）可以由分子的意义中得到可靠的解释。

对于那些直接面对狗而不顾分子生物学的人来说，这一切都顺理成章。爱狗人士滔滔不绝地讨论不同犬种的脾性和心性，仿佛所有这些都与DNA的化学和物理学完全没有什么联系（更别提比DNA更精细的物理层级了），而且仿佛这些讨论单纯发生在“犬种的性格特征”这一抽象的层面之上。而神奇之处在于，爱狗人士用这种方式去谈论和思考也能安然处世，完全不比分子生物学家更差。这种方式是真正有效的！事实的确如此。如果他们（或者分子生物学家）努力以纯粹的物理学或分子生物学的方式去讨论和思考，那么他们将立刻陷入微观实体的细节泥沼之中，而这些构成了狗及其品种（暂不提宇宙的其余部分）的微观实体，彼此交互作用，数量之大难以想象，其细节更是无尽无穷。

上述一切带来的结果就是，讨论狗或猪的最现实的方式，涉及的是高层级的实体，而正如罗杰·斯佩里所说，这些高层级的实体在任意使唤着低层级的实体。还记得吗，真正推倒位于链阵中“质数分支”上的那张坚硬的多米诺骨牌的力量是来自整数641的质数性这一捉摸不定的抽象特质。这正是一种向下的因果性，并直接将我们引向这样一个结论：要想对拥有符号的大脑进行思考，最有效率的方式——而且在大多数情况下，最真实的方式——就是认定是观念和欲望在使唤其内部的微观之物，而并非相反。

---

1. See [Kneebone], [Wilder], and [Nagel and Newman], for reasons to believe strongly in the consistency of *PM*-like systems.

2. For another treatment of the perverse theme of “although” turning into “because”, see Chapter 13 of [Hofstadter 1985].

3. For an amusing interpretation of the infinite repeatability of Gödel’s construction as demonstrating the impossibility of artificial intelligence, see the chapter by J. R. Lucas in [Anderson], which is carefully analyzed (and hopefully refuted) in [DeLong], [Webb], and [Hofstadter 1979].

4. See [DeLong], [Wolf], [Kneebone], and [Wilder].

5. [DeLong], [Goodstein], and [Chaitin] discuss non-Gödelian formulas that are undecidable for Gödelian reasons.

6. See [DeLong], [Boolos and Jeffrey], [Jeffrey], [Goodstein], [Hennie], [Wolf], and [Hofstadter 1979] for discussions of many limitative results such as this one (which is Church's theorem).

7. See [Smullyan 1992] and [Hofstadter 1979] for discussions of Tarski's deep result. In the latter, there is a novel approach to the classical liar paradox ("This sentence is not true") using Tarski's ideas, with the substrate taken to be the human brain instead of an axiomatic system.

8. See [Andersen] for a detailed technical discussion of downward causality. Less technical discussions are found in [Pattee] and [Simon]. See also Chapters 11 and 20 in [Hofstadter and Dennett], and especially the Reflections. [Laughlin] gives fascinating arguments for the thesis that in physics, the macroscopic arena is more fundamental or "deeper" than the microscopic.

9. See [Monod], [Berg and Singer], [Judson], and Chapter 27 of [Hofstadter 1985].

第13章
我的“我”，捉摸不定的心头肉

构成经验的模式

在人类本性的最深处，有一个由熟悉而舒适却又几乎不可能明确定义的抽象模式构成的世界，而我们正漂浮于其中；这些模式包括：“快餐”和“番茄蛤汁”、“黏性”和“古怪”、“圣诞节优惠”和“客服部门”、“竹篮打水”和“我行我素”、“怪咖”和“软肋”、“灌篮”和“底线”、“空话”和“重活”、“阴招”和“食品袋”、“个人音乐会”和“人渣”、“酸葡萄”和“肥皂剧”、“反馈”和“公平竞争”、“目标”和“谎言”、“恐惧”和“梦想”、“她”和“他”——别忘了还有“你”和“我”。

虽然我把上面每一条词汇都放在了引号里，但是我所讲的并不是写在书面的文字，也不是在谈论这些表达所“指向”的在这个世界中可以观察到的现象。我之所言，乃是这些词汇在你我心中所指示的概念——或者换用一个更早的说法，乃是在我们各自大脑中所对应的符号(注)。

我列出这份希望没有那么无聊的小清单（这是我从一个很长的清单中删减而得的），意在努力描述大多数成年人日常心理现实的情况——各种如面包和黄油一样的符号，平时休眠于人的大脑之中，而在一个人展开日常行为时则可能被纷纷唤醒，这些日常行为包括但不仅限于与朋友和同事聊天、停在十字路口等待绿灯、收听电台节目以及在牙医候诊室翻阅杂志。我所列的清单是穿越某种日常的心理空间的一次随机漫步，呈列出来只是为了展现一种有关我们最信任的、深信不疑的那些现象的感觉（“酸葡萄”和“竹篮打水”对我们绝大多数人而言都是相当真实

的感觉），而与这些现象对峙而立的，是令人望而却步的不可触及的夸克和胶子的层面[注]，以及仅仅比其略微容易触及的基因[注]、染色体和转运RNA的层面——我们可能会围绕这些“现实”层面说一些浮皮潦草的空话，但是很少有人会对其进行思考和讨论。

可是，不管我们如何声称这张清单是真实的，它仍然弥漫着模糊、暧昧、令人难以置信而不可捉摸的抽象概念。你能想象对其中任何一个说法进行精确的定义吗？比如，到底什么特质才是“黏性”？你能把它交给你的孩子吗？还有，请给我提供一个识别模式的算法，以便我们可以万无一失地检测出人渣！

以1/2自旋粒子进行自我反思的资本主义单身汉令人摸不到头脑

我们的思考与宏观世界模糊、氤氲的范畴深刻地联系在一起，要想理解这种联系到底有多么深刻，只需要借助一个简单的范例，想一想下面这个奇妙的事实。逻辑学家们的本职就是努力写下精确的逻辑推理法则，赋予语言表达以完璧无瑕的精确性，可他们却几乎从未诉诸粒子和场的层面来支持他们那些根本性的永恒真理的经典范例。相反，他们最常引用的“真理”范例，都典型地运用了范畴模糊的句子——比如“雪是白的”、“水是湿的”、“单身汉是未婚的男子”以及“资本主义在未来几年的美国可能会深陷麻烦，也可能不会”。

如果你认为这些句子确实表达了鲜明的真理，那么请稍停片刻，深思一下……“雪”究竟是什么意思？它是如“将军”或“质数”一样清晰的范畴吗？而“湿”准确地说又究竟意味着什么？它没有一点儿含糊其词之处吗？“未婚”呢？——更不用提“未来几年”和“深陷麻烦”了。模棱两可在此可见一斑！可尽管如此，这种经典的哲学家式的句子，因其处于我们自然漂浮的层面之上，所以在大多数人眼中都比诸如“电子有1/2自

旋”或“电磁感应定律在镜像中是不变的”这样的句子更加符合现实（因而其真实性也更加可靠）。

我们是相对宏观的存在，因此绝大多数人都从不会直接看见或应对电子或电磁感应定律。我们的感知聚焦在更大、更模糊的事物之上。我们内心深处更加信任宏观事物以及表面观察到的因果关系，而不是微观世界的电子作用。我们不断地把这些宏观事物分配给我们高频或低频的心理范畴（高频的范畴如“快餐”和“食品袋”，低频的如“软肋”和“客服部门”）；同时，不管我们观察到的因果性有多么模糊和不可靠，身处这些宏观而模糊的事物之间，它们看起来仍是成立的。

在通常无比混乱的生命世界中，我们对于因果性最为锐利的洞察，无一例外地来自在宏观层面精良打磨的范畴划分。例如，当一位颇具卓识的评论者把一场战争的起源与存在于某些宗教教义之间的古老冲突联系起来时，这场发生在某遥远大陆的神秘战争的原因便有可能突然之间一跃成为我们关注的焦点。另一方面，如果一位物理学家试图用发生在不计其数转瞬即逝的量子力学中亿万次动量守恒的冲撞来解释战争发生的原因，那么完全不会起到任何解释的效果。

我可以如此继续，把同样的话用在恋爱和其他宏大的人生主题之上，说我们总是通过属于宏观世界的捉摸不定的日常模式来感知这些主题，而从不会通过基本粒子的交互作用来看。与其宣称量子电子动力学是“维持世界运行之力”，我宁愿征引如美、慷慨、性、不安全感、忠诚、嫉妒、孤独等等永远捉摸不定的神秘说辞，并确保不会遗漏那种发生在两个灵魂之间而被我们称为“化学反应”的神秘共鸣。法国人对此有一种更妙的说法，即“avoir des atomes croches”，意思是两个灵魂有勾连在一起的原子。

列出这样一份清单虽然有趣，但只不过是一次简单的演练，不会给你带来任何新鲜的启示。我们在此需要领会的要点在于，我们对于生命中每一种事物的感知，从本质上来说，都停留在这个层面之上；而且，

从本质上来说，我们虽然知道我们是由一些肉眼看不见的成分组成的，但是在那个层面上，我们却感知不到任何东西。我承认，少数的例外是存在的，比如我们现代人对于微观的致病源的敏锐意识、我们对于创造生命的精卵结合这一微观事件的兴趣、关于新生儿性别的微观决定因素的常识——这些都是例外中的例外。一般而言，我们就是活在日常概念的世界里，也正是它们定义了我们的现实，而并非微观的事件。

我是一颗奇怪的玻璃球吗？

前文所说的这一切的意思是，我们理解其他生物行为的方式也正是我们理解自身行为的最佳方式[注]——我们依靠的是被称为“希望”、“信念”以及种种名目稳定而又难以触及的内在模式。可是，自我理解的需求却远远超过了这一步。我们受到一股不可抗拒的力量所驱动，创造出这样一个术语，令它获得我们在自己大脑中发现的所有希望、信念和欲望所假定的统一性、内在的连贯性以及时间的稳定性——而那个术语，我们一早就明了于心，那就是“我”。而且，这一走出幕后的高度抽象概念，很快就会给人感觉是全宇宙中最真实的实体。

正如我们深信，引发战争与爱情的是观念与情绪，而不是粒子，我们同样也深信，我们自己的行为是由我们的“我”所引发。在我们的身体内以及属于我们身体的“大指使者”是我们的“我”，是我们在装满了层叠的希望与欲望的箱子里摸到的那颗美妙非凡的玻璃球，而我们可以确凿无疑地感觉到它的形状、硬度与尺寸。

显然，我在此暗指的是“副”——那个信封箱子里不存在的玻璃球。但是，“我”的错觉远比由多层齐整的纸张和胶水制造出来的玻璃球错觉更加微妙和难以克服。这种错觉的坚韧性从何而来？为什么不管遭到多少“硬科学”的攻击，它就是拒绝褪去？为了努力找到这些问题的答案，我下面要把焦点放在制造出一个“我”的那个怪圈之上了——它在何处被

发现、如何生成、如何稳定下来。

我不是一条珍珠项链

首先要说的是，对于我们每个人而言，我们独特“我”的怪圈都安居于我们自己的大脑之内。因此，在每个正常人类的大脑里，都潜伏着这样一个怪圈。实际上，我得撤回这个说法，因为在第15章，我会把这个数字急剧地提高。话虽如此，在最开始的时候，说每个人大脑里仅仅有一个怪圈，这个估算结果也算是差强人意了。

但我提到“一个大脑内的一个怪圈”时，我心中所想的是一个物理结构吗？它是不是某种可见的闭环？或许是由很多神经元首尾串联起来的一个回路？这个中性的环路可以通过一次大脑手术中干净地切除，然后放在桌子上，像一条精美的珍珠项链一样供所有人观赏吗？而那个大脑被“去环路”的人会因此而成为一具丧失意识的行尸走肉吗？

不消说，这几乎与我心中所想毫不沾边。就如同一个音频反馈环并不是一个拥有质量和直径的可触的物体，制造出一个“我”的怪圈也同样并非一个可被触碰或摘除的物体。音频回路可以存在于一个音乐厅“之内”，可它在物理上定位于此的事实并不意味着人们可以捡起它、举起它，更别提测量这种东西的温度和厚度了！“我”的怪圈和一个音频反馈环一样，也是一种抽象物——但是这种抽象物看起来却极其现实，甚至几乎可以与其发生物理上的接触，尤其是对于像我们这些在亨内克量表上读数很高的生命而言。

“我”是我的大脑的最复杂的符号

一个大脑同一个闹壳（以及*PM*）一样，也可以（至少）在两个层

面上接受观察——一个包括了很小的物理过程（可能是包括了粒子，也可能是包括了神经元——随你任选）的低级层面，和一个包括了由感知选择性触发的宏大结构的高级层面。后者在本书中被我称为符号，其在我们的大脑中构成了“我们”的范畴的结构。

在一名正常人类的库存中不计其数的符号之间，有一些远比另一些出现得更加频繁，也更具统治性，而其中之一被多少有点任意地赋予了“我”（至少在汉语中）之名。当我们讨论其他人的时候，我们讨论的是他们的野心、习惯和好恶，相应地，我们就需要为他们中的每一个人构造一个“我”的类比物，而这些“我”自然是栖身于他们的大脑内，而非我们自己的大脑。而这些我们自己的“我”的对应物，必然会被贴上各种各样的标签，视语境而定，可能是“丹妮”、“莫妮卡”，或者“你”、“他”、“她”。

感知一个人的自我与宇宙（当然主要由他的家人、朋友、钟爱的音乐、心爱的书和电影等等组成）的其他部分进行交互的过程，贯穿人的一生。因此，“我”的符号与我们大脑中的所有符号一样，最开始的时候都相当之小，且极为简单，但它不断长大、生长、成长，最终成为我们大脑内部最为重要的抽象结构。但是它在我们大脑中的何处呢？它不在某个很小的定点之上；它四散开来，到处都有它的身影，因为它不得不包罗万象，容纳万千。

内化我们的曾经、我们的意志和我们的将来

我的自我符号跟我的狗的自我符号不一样，虽然断断续续，却可以相当准确地往回探伸到我生命的遥远（并仿佛是无穷无尽的）过往。相比于动物，我们在精密性上完成的这神奇一跃，得益于我们可以无限扩展的人类范畴系统，因为它使我们每一个人都得以构建起我们的情景记忆——那个装满了我们对于事件的回忆的巨大仓库。那些事件或大或

小，或简单或复杂，在过去的几十年时间里发生在我们（以及我们的朋友、家人，以及书中、电影中、报刊文章中的人，以及其他无限多的人）的身上。

无独有偶，在恐惧与梦想的驱使下，我们的自我符号虽然信心不足，可依然不遗余力地窥向我的未来存在，放眼于那一团氤氲的迷雾。我关于自己过往生命的庞大情境记忆，偕同其模糊指向未来之时的对应物（我想我会把它称为我的情境投射），再加入一种奇幻的别样变体(注)或者说无数场景的“虚拟重播”（“要是X发生了，就好了……”；“要是Y从未发生过，那该有多么幸运啊……”；“如果Z能发生，该有多好……”何不将此称为我的情境虚拟呢？）作为装饰，便生成了构成我的“我”的那条无穷无尽的镜壁长廊。

我不能离开我自己而独活

我们观察的不是交互作用的粒子，而是宏观的模式，并有某些特定的东西在这些宏观模式中以某种模糊不清的因果性在使唤着其他东西；我们的身体内以及属于我们身体的“大指使者”是我们的“我”；我们的身体使唤着世界中其余的一切：既然如此，则我们别无选择，只能认定“我”是因果性的主人。在我们每个人看来，这个“我”都是我们所有行动和决策的根本。

当然，这只是真相的一个侧面而已，因为它完全冷落了另一种视角，即有一种非人格化的微观物理学在维持着世界的运转；但是，这种视角的失真扭转却具有令人惊讶的可靠性，并且绝对不可或缺。事实上，在我们由婴儿时代出发，经由童年而进入成人时代的过程中，这种天真的、非物理性的视角的两种特性——可靠性和不可或缺性——将根深蒂固地扎根在我们的信念系统之中。

我还要补充一点，一名粒子物理学家的“我”与一位小说家或鞋店店员的“我”一样根深蒂固，并无差异。对于所有物理性的精深把握丝毫不能撼动几十年来文化与语言的潜移默化，更不用提为其铺平道路的上百万年的人类进化了。“我”的概念既然是一条无与伦比的高效捷径，也必然是一种不可或缺的解释工具；它绝不仅仅是一根可有可无的拐杖，当人们的科学认识足够精密成熟之后，便可以欢欣地丢弃一旁。

自我的缓慢建成

是什么使人类的大脑得以成为可容纳一个自我表征环路[注]的备选之所？为什么一只苍蝇或蚊子的脑不具备同样有效的备选资格呢？就此而论，为什么一个细菌、一个卵子、一个精子、一个病毒、一株番茄植物、一个番茄或一支铅笔不行呢？答案应该是很清楚的：人脑是一个表征系统，而且其范畴在延展性和灵活性上都是没有边界的。对比之下，一只蚊子的脑也是一个微小的表征系统，但实际上并没有包含任何范畴，更不用提灵活性和延展性了。更小的表征系统，如细菌、卵子、精子、植物和温度计等，根本奢谈不到自我表征。而番茄和铅笔更是连表征系统都算不上，所以对它们来说，故事仅此而止。（抱歉了，小番茄！抱歉了，小铅笔！）

所以，人类的大脑是一个拥有丰富的感知反馈潜力的强势备选项，因而也有潜力容纳丰富的自我表征。但是，我们卷入的是一种怎样的感知回路呢？我们在生命的开始，接受的是最基本的关于我们自己的反馈，它们促使我们构建出描述我们最明显的身体部位的范畴，而在这一基座之上，我们很快发展出一套将自己的身体视为灵活的物体的意识。与此同时，我们在因各种各样不同的行为而受到激励或惩罚的过程中，开始发展出一种更加抽象的关于“好”和“坏”以及愧疚与骄傲这些概念的意识，而我们将自己视为有能力通过决策促使事情发生（比如在大腿苦

苦哀求我们走慢一点的时候，我们依然决定跑步冲上斜坡）的抽象意识，也由此开始生根。

尽可能精确地打磨我们发展中的自我符号，对于我们年轻时的生命而言，是至关重要之事。我们想要（需要）在各式各样的社会等级和阶层中找到自己的归属；有时候，即便我们不想知道这些事情，我们迟早还是会明白。例如，我们在很小的时候，都会被人夸奖“可爱”；可是这个信息随着时间的推移，就是会在某些人的身上比在另外的人的身上得到更加牢固的强化。通过这种方式，我们每个人都会意识到，我们到底是“好看”还是“容易上当”，是“厚脸皮”还是“害羞”，是“被宠坏了”还是“有趣”，是“懒”还是“特立独行”，等等。数十种这一类的标签和概念将累加在我们逐渐成长的自我符号之上。

在我们迈过千万个大大小小的经历的过程中，我们对于这些经历的表征也会同样累加在我们的自我符号之上。当然，一段参观大峡谷的记忆不会仅仅累加在我们的自我符号之上，还会累加到我们大脑中很多其他符号之上，但是我们的自我符号确实因为这次累加而变得更加丰富，也更加复杂了。

投出去，再内化反弹

持续不断、坚韧不拔、日复一日，争分夺秒，我的自我符号在时刻经历着塑造与改良——而转过身来，它也触发了大量的外部行为。（或者说，因果性由此表现出来，因为它对世界的感知停留在了这个层面之上，而不是微观的层面。）它眼见自己选择的行动（踢腿、抛起、叫喊、微笑、讥讽、戳刺、旅行、预订、恳求、威胁等等）在它所处的环境中制造出各种各样的东西，做出或大或小的反应，并以其粗线条的范畴（说到它们的粗线条，是别无选择）内化了这些效果。通过像这样的无休无止的随机探索，我的自我符号慢慢地把简洁而有价值的洞见收入

其本性当中，而这个本性嵌于一个博大、多面而部分可预知的世界中的行动选择者与启动者。

更形象一点地说：我把篮球抛向篮筐，要么这次勾手投篮会投中，要么投不中，这取决于在我的手臂、手指、球的旋转、空气阻力、篮筐摩擦力之间发生的一系列我本人意识不到的微观事件。这种伸入世界的微小探索，重复了上百上千次后，能够无比精确地告诉我自己作为一名篮球手的技术水平（并能帮助我判断我是否热爱这项运动）。当然，我对于自己技术水平的感觉只不过是对于有关我的身体和大脑的上百亿个精细事实进行的一种十分粗线条的总结。

同理，我的社会行为也会引发其他智慧生命的反应。这些反应反弹回我这里，而我利用我自己的符号集合对其进行感知，以此经由我对其他人发生的影响而间接地感知自己。我从他人的眼中建立起关于我是谁的认识。我的自我符号是从元始的虚空中集结而出。

像卡西迪牛仔一样微笑

大概在我六岁时的一天早晨，我鼓足了所有勇气，在一年级的展示和讲述课上站了起来，骄傲地宣称："我可以做出跟卡西迪牛仔（Hopalong Cassidy）一模一样的微笑。"（我不记得我是怎样说服我相信自己具备这项伟大能力的了，但是我对此确实深信不疑。）然后，我就在全班同学的面前绽放了自己精心练习过的笑容。几十年过去了，在我的情景记忆中，这次英勇的表现依然历历在目。不过，令人遗憾的是，对于我当时敬爱的老师、温柔甜美的麦克马洪小姐以及同班小同学们的反应，我只留下了最模糊的印象。可不管他们的反应具体是怎样的，却都毫无疑问地对我早期生命的定型产生了影响，因而也影响了我那逐渐成长、缓慢稳定下来的"我"。

我们所做之事——我们的“我”告诉我们去做的事——都有后续的结果，有的是正面的，有的是负面的。在岁月推移之间，我们会努力塑造我们的“我”，令其不再指引我们走向负面的结果，而更多地指向正面的结果。我们会看到我们的卡西迪牛仔微笑到底是一次大的成功还是一次大的失败，而只有在前一种情况下，我们才愿意再把它拿出来炫耀。（说实话，从一年级以后，我就再也没有把它拿出来现眼了。）

当我们年纪稍长以后，我们会观察自己的俏皮话到底是波澜不惊，还是会激起赞赏的笑声，而根据观察的结果，我们或改进我们编俏皮话的风格，或学会进行更严格的自我审查，或者双管齐下。我们还会尝试各种各样不同风格的服装，然后从其他人反馈的字里行间，解读出我们穿什么样的衣服好看，穿什么样的衣服不好看。当我们因为小小的谎言而受到责备时，我们或者决心不再说谎，或者学会更巧妙地说谎，同时，我们也把关于我们有多么诚实的知识纳入到我们的自我符号之中。说谎的情况，当然也同样适用于吹牛。我们大多数人都会锻炼我们对于语言的使用，令其可以适应各种不同的社会规范，有时候我们是有意在做这件事，有时候没那么用心。复杂性的等级是无限的。

关于我们的我的谎言

有一种奇怪的隐藏结构深植于我们每个人最为核心之处。一个多世纪以来，临床心理学家们都在努力试图理解这种结构的本性，而其中一部分人也写下了颇具洞见的著作。几十年前，我读到了精神分析学家卡伦·霍妮（Karen Horney）所著的几本书，它们给我留下了持久的强烈印象。例如，霍妮在《我们内心的冲突》（*Our Inner Conflict*）一书中说到了一个人从自我中塑造的“理想化意象”。虽然她主要的焦点在于我们是如何受到神经官能症的折磨，但是她笔下所言却具有更为广泛的适用性。

……它（理想化意象）表征了一种艺术创造物，对立项似乎在其中达成了和解……

理想化意象也许可以被称为一种虚构的或幻想的自我，但那只说明了真相的一半，所以容易引人误解。在其创造物中运转的痴心妄想自然令人咋舌，特别是因为，它发生在了一个原本坚守现实基础的人身上。但是这并不能表示，它完全是虚构的。它是一个融合了最为现实的因素并由其决定的想象创作物。它通常包括了那个人最真实的理想的痕迹。虽然宏伟的成就是错觉，但是这些表面成就之下作为支撑的潜能却常常是真实的。更重要的是，它发自非常真实的内在需求，履行了最为真实的功能，并且对它的创造者产生了真实的影响。在其创造物中运转的过程遵循了确定的法则，类似地，我们可以经由关于其个别特征的知识而准确推断出特定之人的真实的性格结构。

霍妮所说的显然不是一个人对其表面特征感知的意识，比如身高和头发的颜色，也不是一个人关于轻度抽象物的知识，比如一个人的工作是什么及其是否热爱这份工作；她所说的，是一个人用一生的时间形成的那个（多少有些扭曲的）意象，其描绘的是自己最深刻的性格特质、是自己在各种模棱两可的社会等级中的定位、是自己最大的成就和失败、是自己得偿之所愿和未尝之夙愿，不一而足。她在书中强调的是这个意象中的错觉方面，因而也是趋于造成伤害的那些方面，但是这种神经官能症的扭曲所处的那个完整的结构却远大于此。这个结构就是我在此称为“自我符号”之物，或者还有一个更简单的说法，就是“我”。

霍妮更早的一本书，《自我分析》（*Self-Analysis*），致力于处理当一个人努力改变自己的神经官能症倾向时将遭遇的复杂挑战，而它不可避免地关注了一个相当具有悖论性的观念，即自我探索自身、有意地试图左右自己深处的改变。这里不太方便深入探究这些错综复杂的议题，但是我还是匆匆提了一嘴，因为这也许帮助我提醒读者：在所有人类生命的核心之处，都存在着弥天的心理复杂性。

环路模式“我”的锁定

下面，请让我用一种更加抽象一点的说法总结上文要义。大量被我们称为“我”的东西，在某个特定的时间，共同引发了某个特定的外部动作，就像一块投入水中的石头，引发了层层扩展的涟漪。很快，我们的行动带来的数不清的后果开始向我们反弹回来，就像第一拨涟漪触到池塘岸边后又反弹回来。我们回收到的东西给我们提供了一个机会，可以感知我们逐渐变化成形的“我”的所作所为。上百万的微小反射信号从外部冲击着我们，不论是在视觉上、听觉上、触觉上或者是任何感知层面，当它们在我们的大脑内落地时，都能激发起二级、三级的内部波纹。最终，从这一簇凌乱的符号中筛选出一小把激活的符号——这很小的一套精挑细选出来的范畴，便构成了对于我们刚刚所做之事的粗线条的理解。（例如，“投篮——我的勾手就偏出了一丁点儿！”或者也许是“哇，我的新发型把他的魂儿都勾住了！”）

因此，当下的“我”——那一套最新升级的回忆、心愿、热情和困惑——通过摆弄事物与他人那片浩瀚而不可预知的世界，激起了某些快速的反馈，而这些反馈一旦以符号激活的形式被吸收，就会引发“我”的一次极小的修订；由此一轮接着一轮，一刻接着一刻，日日不息，年年不止。作为我们最内部的实质而发挥作用的这一抽象结构，正是以此方式，经由符号激发行动并反过来触发符号的环路，才得以缓慢而笃定地发生进化，并由此把自己比以往更严格地锁定在了我们的思维之中。的确，就像一个音频反馈环发出的一声尖响将不可避免地调整自身并最终稳定在系统的自然共振频率之上，经年之后，“我”也将不可避免地让自己逐渐汇拢并稳定起来。(注)

我不是一个视频反馈环

类比时间又到了！我想再一次启用那个视频反馈环的世界，因为我们现在讨论的东西中有很大一部分都可以在这个简单得多的界域里找到与其对应之物。一起事件发生在摄像头之前，便会被传送到屏幕之上，但屏幕上呈现出来的只是一个简化的形式，因为原本连续的形状（纹路极为精细的形状）被转移到了一张布满独立像素的格子（一种粗纹理的介质）之上。然后，这个新的屏幕又被摄像头采集到，并反送回去，如此往复不停。所有这一切的结果就是，屏幕上出现了一个很容易感知到的格式塔图形——某种保持稳定却又是独一无二而前所未见的涡旋。

由此可见，制造出一个人类的“我”的，是一个怪圈，其与视频反馈环之间存在着一个关键性的差别。在电视设备中，如我们之前所见，每一阶段的所有的感知都发生在环路之外——这个环路只不过是单纯的像素的传递和接收而已。电视的环路并不是一个怪圈——它只是一个反馈环。

相比之下，在任何一个引发了人类自我的怪圈中，于感知、抽象化以及范畴划分等不同等级之间来回转换的行为是位于其核心的、不可或缺的要素。正是从原始刺激到符号的向上跃迁，为这个环路注入了“怪性”。一个人自我的整体性格式塔式“图形”——由怪圈构成一个人的“我”的“稳定涡旋”——并不是由一个事不关己的中立摄像头所接收到的，而是以一种高度主观性的方式被感知到的，经历了范畴划分、心理回放、反思、比较、反事实推理和决断等具有主动性的过程。

“我”被根深蒂固地扎稳了根……

当读到我在一年级的展示和讲述环节中尝试表演卡西迪牛仔式微笑的壮举时，你的心里也许会划过这样一个问题：“侯世达怎么又把粒子抛在脑后了？”当然，在听到这则逸事时，你也有可能并没有想到这个问题。我真心希望是第二种情况！说实在的，任何一个头脑清醒的人

（包括最顽固的粒子物理学家）在读到那一段话的时候，怎么可能会产生那样一种古怪的想法呢？在那个语境中，即便最模糊的、最浮皮潦草地提及粒子物理学，看上去都像是引入了一个荒谬的无理推论（nonsequitur）。胶子、μ介子、质子和光子这些东西，跟一个小男孩对自己偶像卡西迪牛仔的模仿，究竟能有什么关系呢？

虽然可以肯定的是，在那个小男孩的大脑里，一定有大量的粒子在“最底层”持续不断地翻滚蒸腾，可是它们同一个闹壳内无数个来回冲撞的小磁交弹一样是隐而不见的。罗杰·斯佩里（这是我后来的一个偶像，要是我一年级的时候就能读到并理解他的作品，我很可能会鼓起勇气站起来向我的同学们宣称：“我可以像罗杰·斯佩里一样进行哲学思考！”）会格外指出，这个小男孩脑子里的粒子只是在为等级高得多的符号化事件服务（即受到后者的差使），后者才是男孩的“我”参与的事件，而且他的“我”还同时在其中成形。当那个“我”发展出更高的复杂性，并在自己眼中变得无比现实的时候（即对于男孩把生命中从未重现的事件范畴化并试图理解的努力是最不可或缺的依靠），另一种与其竞争的“无我”的理解世界的方式，出现的可能性就变得微乎其微了。

不仅我本人越来越习惯“我”这个东西对我所做之事负责这一事实，与此同时，我的父母和朋友也越来越确信“在那里”确实有某种看起来非常现实的东西（换句话说，某种很像玻璃球的、具有独特的“硬度”、“韧性”和“形状”烙印的东西），值得被称为“你”、“他”或“道格”，或者被道格称作“我”——所以，再一次，这个“我”的现实感又被千变万化的方式反反复复地加强了。等到这个大脑已经在这具身体里存活了一两年左右之后，“我”的概念就被锁定其中，再也没有一丝反转的希望了。

……可我是真实的吗？

话虽如此，不管这个“我”具有多么不可撼动的稳定性和显明的实用性，它真的是一个真实的东西吗？还是说，它仅仅是一个令人感到舒适的迷思？我认为，我们在这里有必要求助于某些古色古香的怀旧式类比了。所以我问你，亲爱的读者，温度和压力到底是真实的东西，还只是约定俗成的说法（facons de parler）？彩虹是一种真实的东西，还是不存在之物？或许更为切近的一个类比是，我在装信封的箱子里发现的“玻璃球”是真实的吗？

要是那个箱子当时被封住了，而我没办法逐一查看那些信封，会怎么样？要是我在形成对那个信封箱的认识时，只把其中上百张信封作为一个单一的整体来处理，于是不再可能在粗略与精细的视角之间来回转换，又会怎么样？要是我之前根本就不知道箱子里有信封，只是单纯地以为里面有一种多少可以挤压并有点柔韧性的软软的填充物，而我可以用我的整个手掌抓住它，并感觉到在那堆软软的物质的中心有一种更坚硬的、显然是球形的东西，那又会怎么样？

如果，除此之外，谈论这个假想中的玻璃球被证明在我们的生活中具有相当有效的解释力，又如果，在此之上，我的所有朋友都拥有相似的纸箱，而且每个人都在不停地——且毫无疑心地——谈论他们的箱子里的“玻璃球”，那么我很快也就会相当难以抗拒地接受我自己的玻璃球作为世界一部分的存在，并在我对于这世界中发生的多种多样现象的解释中频繁地加以征用。事实就是如此，如果有任何不合群的怪人拒绝承认自己的纸箱里存在玻璃球，人们就会指责说，是他自己弄丢了自己的玻璃球。

同样的道理也适用于“我”的概念。因为它囊括了我们在这个世界上感知到的因果性的真正重要的方面，并且如此整洁与高效，所以我们没办法不为我们的“我”以及其他人的“我”赋予真实性[注]——实际上是在所有的可能性中位于最高级别的真实性。

构成一个自我的怪圈的尺寸

让我们再一次回过头，来讨论一下蚊子和狗。它们拥有任何像“我”的符号这样的东西吗？在第1章中，当我说到“小灵魂”和“大灵魂”时，我曾说过，这并不是非黑即白的事，而是程度上的差别。因此，我们不得不把问题改成，在一只蚊子的脑袋瓜里，是否有一个怪圈——精密复杂、等级交错的反馈环？一只蚊子是否具有关于自我的丰富的符号表征，包括对于它的欲望的表征和威胁这些欲望的事物的表征，以及是否具有一个通过与其他自我相比较而得出的自我表征？一只蚊子能够产生哪怕只有一点点近似于“我能做出跟卡西迪牛仔一模一样的微笑”这样的想法吗？——例如，“我能像嗡嗡叫的贝蒂一样咬你！”？我想，这些问题以及其他类似问题的答案是很明显的，“绝无可能！”（这都怪一只蚊子脑中那简单到不可思议的符号集合，这个集合的大小甚至跟抽水马桶和温度计的符号集合相差无几。）据此，我可以毫无顾忌地打消念头了：在一只蚊子那样微小而孱弱的脑中，不可能存在属于自我的怪圈。

另一方面，当我们把目光转向狗时，我不出意料地发现，我们有更多的理由可以认为其中至少存在这种怪圈的雏形。狗不仅拥有容纳了很多微妙范畴（比如“UPS快递卡车”或者“我可以在房子里叼起并含在嘴里走来走去而不会被惩罚的东西”）的大脑，而且似乎还对于它们自己和他人的欲望有某种初步的理解，不管那个他人是人类还是其他的狗。当一条狗的主人对它不满意的时候，它常常是明白的，并会摇尾乞怜，希望恢复良好的关系。可尽管如此，一条狗还是明显缺乏一个可以任意扩展的概念库，因而只具备一种最初级的情景记忆（当然，它还完全欠缺任何长久的储存沿着心理时间线串成的想象中未来事件的故事仓库，更不用说盘旋在过去、现在乃至未来的反事实景象了），其自我表征也就必然比成年人的自我表征简单得多。而正因如此，我们才说，狗拥有一个比人更小的灵魂。

机器人汽车的所谓自我

当读到有关“斯坦利”（Stanley）的消息时，我被深深地震撼到了[注]。它是斯坦福人工智能实验室开发的一台机器人汽车，刚刚在不久之前，仅凭它的激光测距仪、电视摄像头和GPS导航系统，独自横跨了内华达沙漠。我禁不住问自己：“斯坦利拥有一个多大的‘我’呢？”

在这场大获成功的沙漠穿越之后，一位雄心勃勃的工程师随即接受了采访，这位因特尔公司（你要记住，是因特尔公司制造了斯坦利车上的电脑硬件）的研发总监大言不惭地宣称：“深蓝[注]只具有处理能力。它不会思考。斯坦利是会思考的。”

好吧，对于斯坦利所代表的令人瞩目的集体成就，我是充满敬意的，但这并不能阻止我把他的发言评价为恬不知耻、彻头彻尾而且天真幼稚的宣传炒作。我眼中所见的与他所说的完全不是一回事儿。不论何时，如果斯坦利能够获得像滚雪球一样地形成如本章开篇所列的那些无限范畴的能力，那么我才会欣然地附和说斯坦利是会思考的。但是就目前而言，它可以毫发无损地横穿沙漠的能力，在我看来，完全可以类比于蚂蚁循着浓厚的信息素轨迹而活着穿过一片空地的能力。一台机器人汽车具有如此这般的自主性，无论如何都不算是一件可以让人嗤之以鼻之事，但是它距离思考、距离拥有一个“我”，依然遥不可及。

在驾驶过程中的某一点，斯坦利的视频摄像头拍到了在它前方的另一辆机器车（那时卡内基·梅隆大学研发的竞争车辆H1），而斯坦利最终超了H1，从此一骑绝尘。[注]人们可以在摄像头内置的录像带里看到这件事的发生，而这也成为这个故事的高潮。在这个关键的时刻，斯坦利有没有认识到另一辆车也是“像我一样”？当它欣然超过H1而扬长而去时，斯坦利有没有想到“上帝赐予我奔跑的力量”或者“哈哈，超过你了！”？这么一说，我为什么要写，斯坦利“欣然超过”H1呢？

让一个机器人拥有这样的想法或者感觉还需要什么？让斯坦利那个锁定位置的电视摄像头能够调转过来拍到它自己，并让斯坦利得以收到关于自己的视觉信号，这样就足够了吗？当然不够。或许，这是在获得一个“我”的漫长过程中不可或缺的一步，可是想一想我们有关小鸡和蟑螂的知识就能知道，对于一个身体部位的感知并不能创造一个自我。

一个反事实的斯坦利

斯坦利缺少一种能赋予它一个“我”的东西，而其似乎也并非自行驾驶汽车开发者研究计划当中的一部分：这就是对于它在世界中所处位置的深度理解。当然，我这么说的意思并不是指这辆汽车在地球表面的定位，这可以由GPS精确测量到厘米的级别；我指的是一种对于这辆汽车自己的行动以及它与其他汽车关系的丰富表征，即对于其目标和“希望”的丰富表征。这就要求这辆汽车具备关于它经历过的成百上千次经历的完整的情景记忆，具有一个情景的投射轨迹（它预料在它的“生命”中会发生什么、它希望发生什么以及它害怕发生什么），并拥有一个情景的虚拟状态，详细地展现出它关于自己擦肩而过的险境的想法，以及假如事情换一种样子发生，最有可能出现的情况。

这样一来，机器车斯坦利就一定能在自己的心里打起下面这种假设性未来的念头：“噫，我真担心，H1会不会故意在我的前面来回变道，不让我超车，或者甚至把我挤到马路外面，把我撞倒那边的阴沟里去！如果<u>我</u>是H1的话，<u>我反正是</u>会这么做的！”然后，过了一会儿，它一定能乐此不疲地玩味如下这种反事实的想法：“吁！我好开心，还好H1没有我担心的那么聪明——也可能H1本来就没有我这么强的竞争力！”

《连线》（*Wired*）杂志中的一篇文章描述了斯坦福开发小组一次近乎惊慌失措的经历。那是在沙漠挑战迫在眉睫的时候，他们忽然意识到还缺少一样东西。文章信笔陈述“他们需要一种自我意识的等价算

法”，随后便继而声称，他们很快就真正实现了这个目标（花了他们整整三个月的工夫）。还是那句老话，这个团队的伟大成绩值得所有人脱帽致敬，可是人们还是该清醒地认识到，在斯坦利的内部，没有任何一样东西能配得上“自我意识”这一高度拟人化的标签。

斯坦利机械计算装置内部的反馈环，足以成功地指引它行驶在一条绵长的尘土路上，绕过地上布满的坑坑洼洼，穿过锯齿状的仙人掌和风滚草而毫发无损。我向它致敬！但是如果人们把眼光从单纯的驾驶上移开，放到思维和意识之上，那么斯坦利的反馈环还是不够怪——连接近都算不上。人类要想借助集体的力量创造出一种人造的“我”，还有相当长的一段路要走。



1. See [Hofstadter 1979], especially the dialogue “Prelude… Ant Fugue” and Chapters 11 and 12, for a careful discussion of this notion.

2. See [Weinberg 1992] and [Pais 1986] for attempts at explanations of these incredibly abstruse notions.

3. See [Monod], [Berg and Singer], [Judson], and Chapter 27 (“The Genetic Code: Arbitrary?”) in [Hofstadter 1985].

4. See [Dennett 1987] and [Dennett 1998].

5. [Steiner 1975] has a rich and provocative discussion of “alternity”, and the dialogue “Contrafactus” in [Hofstadter 1979] features an amusing scenario involving “subjunctive instant replays”. See also [Kahneman and Miller] and Chapter 12 of [Hofstadter 1985] for further

musings on the incessantly flickering presence of counterfactuals in the subconscious human mind. [Hofstadter and FARG] describes a family of computational models of human thought processes in which making constant forays into alternity is a key architectural feature.

6. See [Morden], [Kent], and [Metzinger].

7. See [Dennett 1992].

8. See [Kent], [Dennett 1992], [Brinck], [Metzinger], [Perry], and [Hofstadter and Dennett].

9. See [Davis 2006].

10. 由IMB研发的国际象棋机器，曾在1997年击败世界冠军加里·卡斯帕罗夫（Gary Kasparov）。

11. 顺带一提，我在上面的文字中小心翼翼地回避了代词“他”，但是在涉及斯坦利的新闻报道中，这种用法却是司空见惯，而且考虑到他们给这辆车起了一个人类的名字，所以在人工智能实验里大概也是如此。令人遗憾的是，这种语言学上的草率处理开启了一股不良之风，很快就吹向了完全的拟人化。

第14章

观者的“我”中的怪性

我们头脑中的惰性海绵

读完前面的章节后，你可能会纳闷，为什么我要把贯穿人类一生的自我表征环路称作一个怪圈？你做出决定、采取行动、影响世界、接收反馈、将其融入你的自我，然后经过更新的“你”继续做出决定，如此往复，周而复始。毫无疑问，这形成了一个环路——可是，我一直所说的构成怪圈的必要元素，即那种悖论性的特质，到底何在呢？这为什么不是一个普通的反馈环呢？这种环路模式与库尔特·哥德尔在《数学原理》中意外发现的那个典范性的怪圈，有什么共同点呢？

首先，大脑似乎先天上便不适合作为自我指涉及其反直觉的丰富后果的基底，就如同《数学原理》极其严苛地将自我指涉驱逐在外。一个人类的大脑只不过是一个巨大的紧紧楔入硬邦邦的头盖骨里的毫无生气的分子海绵球(注)，就如同一块榆木疙瘩一样木讷地杵在那里。自我指涉及自我为什么偏偏要潜伏在这样一个奇怪的载体之中呢？这跟潜伏在一块花岗岩里有什么不同吗？大脑里的“我”到底在哪里？

在《数学原理》那座磐石要塞中必然会发生某些特别奇怪的事，使类似“我是不可证的”那种哥德尔式句子的非法的“我”得以潜入进来；无独有偶，在那个塞满了无生命分子的骨制脑壳中也必须发生某些特别奇怪的事，才能引发一个灵魂、一丝“光亮”、一种独特的作为人的本质、一个“我”。而且别忘了，并不是所有的大脑都会拜某种“正确的材料”(注)所赐而魔法般地迸出一个“我”来；只有当那个载体中出现了合适的模式

时，这一切才会发生。没有这种模式，这个系统就跟它表面看起来的样子没有什么不同了：它只是一坨海绵状的物质，没有灵魂，没有“我”，没有任何内在之光（inner light）。

喷射化学物质

当生物脑最早出现时，它们还只是微不足道的反馈装置而已，其精密性还不如抽水马桶的浮球机械装置或者挂在你家墙上的温度计，而跟这些设备一样，它们也在有选择地推动原始的有机体获取某些特定的东西（食物），同时远离另一些东西（危险）。然而，进化的压力逐渐让大脑对于环境的治疗类选法[注]越来越复杂，越来越层级化，而最终（我们这里说的是几百万乃至几十亿年之后），接收反应的范畴集合扩大到极为丰富的水平，使得整个系统可以像一台线缆足够长的电视摄像头一样，有能力在某种程度上“指向”自身。这丝最早的自我的微弱闪光，形成了意识与“我”的胚芽，可是余下的谜团依然巨大。

不管大脑变得多么复杂和精密，在根底处，它们依然永远都只是一套细胞集合而已，在彼此之间来来回回地“喷射化学物质”（这个说法借用自先锋机器人学家和煽动型作家汉斯·莫拉维克[注]），这颇像一座巨大的炼油厂，液体在罐子间永不停歇地抽来抽去。一个抽泵液体的系统怎么可能容下一种上下颠倒的因果性，允许意义比物体及其运动拥有无限高的重要性呢？喜悦、悲伤、对印象派绘画的热爱以及调皮的幽默感，怎么可能在这样一个冰冷而无生命的系统中安身呢？人们要想找一个“我”，还不如去寻访一座磐石要塞、一个抽水马桶的水缸、一卷卫生纸、一台电视机、一支温度计、一枚跟踪导弹、一堆啤酒罐或者一座炼油厂。

有些哲学家认为，我们的内在之光、我们的“我”、我们的人性、我们的灵魂，都是从那个基座自身的本性中散发出来的——也就是说，来

自碳元素的有机化学性质(注)。我觉得这种看法就像一棵无比怪诞的树，上面悬挂着花里胡哨的意识吊件。基本上，这就是一句什么都解释不了的故弄玄虚之谈。碳元素的化学性质凭什么就拥有与其他所有的化学物质都完全不同的魔法属性呢？而且那个魔法属性到底是什么？它如何让我们成为有意识的生命？如果是有机化学决定了一切，为什么单单只有脑子是有意识的，而不是膝盖骨或者肾脏？为什么和我们一样也是以碳元素为基础的蚊子，却没有跟我们一样有意识？为什么奶牛没有跟我们一样的意识？组织形式或模式在这里一点作用都没起吗？它当然起作用了。而如果它真的起作用了，那么为何不能是它扛起了全部的重任呢？

哲学家们声称，碳元素的化学性质具有某种对于意识不可或缺的难以言喻的奇妙性质，由此以关注载体而忽视信息、关注陶罐而忽视图案、关注字体而忽视故事本身的方式而错失了真理。丹尼尔·丹尼特（Daniel Dennett）曾在驳斥约翰·塞尔令人生厌的“正确的材料”时机智地评论道：“无关肉体，重在动作。”（It ain’t the meat, it’s the motion.）依我所见，只有当你知道如何看待脑中发生的动作时，肉体中发生的魔法才能真正算数。

符号的庄严舞步

如果你不再把焦点聚集在脑中喷射的化学物质之上，而是把观察的等级向上迁移，并把那个低的层级抛在脑后，那么大脑就完全呈现出另一番面貌了。而我之所以构想出闹壳的隐喻，目的正是为了让我们能更加方便地讨论这种向上的跳跃；所以容我再强调一下它的关键意象。通过把视角从疯狂冲撞的小磁交弹层面向上拉远，转而在一个加速的时间轴上观看这一系统，人们便开始看到其他实体出现在视野中心；当小磁交弹原本混乱的翻滚蒸腾变成一团模糊雾影时，之前完全不可见的实体便现身了。正是在这个层面上，说也奇怪，意义诞生了。

充满意义的小磁交弹球映入眼帘，在一片模糊的混汤中扬起庄严的舞步；它们从不曾有片刻闪念疑心这片混汤是由名为“小磁交弹”的相互作用的磁性小弹珠构成的。而我之所以说小磁交弹球“充满了意义”，当然不是因为它们渗出了某种被称为“意义”的神秘的黏糊糊的语义汁液（虽然某些痴迷于肉体的哲学家可能会产生这个想法），而是因为它们庄严的舞步与它们周遭世界中发生的事件形成了深度的同步。

小磁交弹球与外在世界同步的方式，正是《面包师的老婆》中小美的回家与迷途妻子奥蕾丽的回家发生同步的方式：情境A与情境P之间存在多重面向的切合。然而，电影高潮处的两种情境的切合，只不过是编剧生造的一个玩笑；任何一位《面包师的老婆》的观众都不会假定，小猫的恶作剧与妻子的恶作剧将在接下来几个月的时间里保持吻合。我们都知道那不过是一次巧合，所以才会感到它相当幽默。

相比而言，在闹壳中起舞的小磁交弹球将持续地追踪世界，*将*与它同进同退，*将*同它保持一致。这（由作者钦定！）正是一个闹壳的终极本性。小磁交弹球与外部世界中发生之事*系统化*地保持统一步调，就像在哥德尔的建构中，质雅数与*PM*的可证公式系统化地步调统一。这是我们说小磁交弹球拥有意义的唯一理由。不论以何物为基底（substrate）——组装玩具、厕纸、啤酒罐、小磁交弹、自然数、神经元，意义都是由可靠而稳定的契合（alignment）所带来的一种自动化的、不可避免的产物；这是我们在第11章中学到过的道理。

我们自己的大脑跟闹壳并没有什么区别；当然，还有唯一的一点区别，就是闹壳只是我一次小小的异想天开，而人类的大脑则不是。我们大脑中的符号真的*在*施展那些它们得心应手的魔法，而且是在神经事件的电子化学混汤之中施展。从最初的原始脑进化出我们的大脑，花了数百万年的光景，可奇怪的是，意义竟然就那么悄无声息地溜了出来，几乎无从察觉。事实是，并没有某个人曾设计过一个宏大的计划，提前几百万年便预定了将有高层级的意义结构——表征抽象范畴的物理模式

——在某一天入驻一个大大的高级的大脑；恰恰相反，这种模式（本书的“符号”）的出现毫无计划可言，只是当越变越大的大脑在帮助生命在残酷凶险的世界中越来越好的存活时，这种极为有效的进化方式便催生了这么一种副产物。

当高级的哥德尔式的意义攻其不备地现身于《数学原理》这座由罗素加以终极防御的碉堡中心时，罗素遭到了出其不意的重击；同理，如果除了汉斯·莫拉维克的喷射化学物质的层面之外，一个人从未想到过还可以在任何其他的层面上看待大脑，那么当他目睹符号的出现时，也会感到极其的震惊。正如哥德尔看到了把注意力转换到*PM*串的完全不同的层级之上后将收获的巨大潜能，我也建议（虽然我远非第一个如此建议之人）我们一定要把注意力转移到大脑活动的更高层面之上，由此去发现符号、概念、意义、欲望，以及最终发现我们自己。

有趣的是，我们人类都已经把焦点集中在那个层面了，根本没有任何其他的选择。我们无意识地把我们的大脑行为作为完全的符号化行为来看待。我觉得这件事存在某种奇妙的怪异性和颠倒性，下面我就将利用一则寓言来努力揭示出其中的原因。

阿尔弗伯特参观奥光星球

如果你愿意的话，请想象一个小小的、孤单的奥光星球（Austrainius），星球上唯一的居住者是名为“库德尔人”（Klüdgerot）的部落。从无法追忆的时代开始，库德尔人的奇妙生活就发生在长满极长的*PM*串的浓密森林中。森林中有些*PM*串是可以供他们安全食用的（这是他们唯一的营养来源），而另外的一些*PM*串则是他们吃不得的，因为这有致命的毒性。幸运的是，聪明机智的库德尔人发现了一种区分两种对立类型的*PM*串的方法。某些特定的串，当接受视觉的审查时，会形成一个信息，并以轻快的库德尔语说出：“我是可以食用

的。”而其他的串形成的信息，用库德尔语说就是：“我是不可食用的。”而相当神奇的是，在哥神（Göd）慈悲之力的沐浴之下，每一个声明自己可以食用的*PM*串都真的是可以食用的，而每一个声明自己不可食用的*PM*串实际上也都是不可食用的。由此，库德尔人在他们丰产的星球上兴盛繁衍了无数个尔年（öörs）。

在奥光纪元斯波尔月（möönth of Spöö）一个命运转折的尔天（döö），一艘外观奇怪的橘色太空飞船从天呼啸而降，它来自遥远的联合星球（Ukia），精确地降落在奥光星球的北尔极点上。从飞船里走出一个体型庞大笨重的白头（white head）外星人，自我宣告说：“看好了。我是阿尔弗伯特。”话音未落，外星人已经踏步踩进了奥光的丛林之中，它不仅在那里度过了斯波尔月余下的日子，而且还度过了整个布鲁尔月（Blöö）。然后，它拖着沉重的脚步走回飞船的降落点，除了衣服脏旧了一点之外，没什么大碍。第二尔天，在明媚的大清早，阿尔弗伯特便庄重地召集了奥光星球上所有的库德尔人前来赴会。他们刚刚聚齐，阿尔弗伯特就迫不及待地开口了。

“尔天好，善良的库德尔人们，”阿尔弗伯特拖着长腔展开吟诵。“我很荣幸地向你们汇报，我已经做出了震动奥光的科学发现。”库德尔人就算心存怀疑，也都在不失尊重地安静聆听。“生长在这个星球上的每个*PM*串，”阿尔弗伯特继续说，“并不仅仅是修长曼妙的藤蔓，而且令人吃惊的是，它们还是可以被阅读和理解的信息。别不信我！”听到这番毫无新意的发言，很多库德尔人不约而同地打起了哈欠，而且有人喊道：“给我们讲讲吧，大白头！”这引发了窸窸窣窣的阵阵笑声。阿尔弗伯特受到了鼓励，于是真的给他们讲了起来。“我有个奇妙的发现。每个*PM*串都在用我美丽的母语，阿尔弗伯特语，做出关于某种已知为‘自然数’的神奇实体的声明。可想而知，你们很多人都已经迫不及待地想听我用你们可以理解的简单说法来解释清楚这些所谓的‘自然数’到底是什么了。”

听到这里，聚集的人群中发出了一阵巨大的沙沙的噪声。阿尔弗伯特不知道，库德尔人在过去无数个世代中，已经认定那种被称作"自然数"的实体是不可理解的抽象物了；实际上，自然数早在很久以前便遭到了众人的唾弃，连同它们所有的名字一起，成为这个星球上永久的禁忌。显而易见，阿尔弗伯特带来的消息在这里是不受欢迎的。它当然是错的（这自不必说），但是它不仅仅是错了而已；它完全是荒诞不经的，除此之外，还令人作呕。

然而，白头的阿尔弗伯特并没有意识到自己的言论所搅起的怨愤。不顾台下暴众发出的越来越不耐烦的沙沙声，他继续说道："没错，奥光星球的居民们，虽然这听起来实在美妙得难以置信，但是每个*PM*串中都包含有意义。我们只需要找到解读它们的正确方式，通过建立合适的映射关系，一个人便能够……"

突然之间，骚乱爆发了：谁叫阿尔弗伯特刚刚说出了那个下贱的"一"字？那可是所有自然数中最令人恐惧的一个名字，在很久以前就已经被禁绝了。"赶走那个外星人！砍掉它的大白头！"激愤的暴众大喊。片刻之后，库德尔人蜂拥而上，那个出言不逊的外星人被裹挟在拥挤的人群之中。然而，即使当它被众人拖拉驱赶的时候，自以为是的阿尔弗伯特仍在坚持向库德尔人不厌其烦地诉说，自己是在努力启迪他们的民智，通过用他们不了解的语言解读那些串，它可以观察到他们看不见的重大事实，而且……然后，愤怒的人潮早把阿尔弗伯特的宏伟言论彻底淹没了。

正当这个厚脸皮的外星人即将面临可怕的厄运时，在库德尔人之间却突然爆发了一阵骚乱；他们竟然差点儿忘记了举办一场"可怕厄运执行前夕宴会"，那可是库德尔人古老而备受尊崇的传统！一个小分队被派到哇尔主行星公园去采集最香甜可口的*PM*串，那里是从未有库德尔人敢于涉足的圣地；小分队满载而归，从哇尔带回了饱满多汁的*PM*串，每一串都清楚地声明："我是可以食用的。"小分队赢得了雷鸣般的

掌声。当库德尔人表达完对于哥神的感恩之后，传统的“可怕厄运执行前夕宴会”就开始了。这时，阿尔弗伯特终于意识到，自己距离可怕的厄运已经近在咫尺了。当这一凶险的事实占据了它的头脑之后，它的大白头开始飞速地旋转起来，然后挣扎凫水，然后……

宽宏大度的阿尔弗伯特仍然在试图尝试拯救那些被蒙在鼓里的库德尔人，他大叫道：“听我说，求求你们了，我的朋友们！你们采到的那些*PM*串是有毒的！愚蠢的迷信骗你们相信它们有丰富的营养，但这不是真相。当作为信息编码之后，这些串给出的有关自然数的陈述全都是虚假的，没有人，我要再说一遍，没有任何人！可以吞食它们。”但是这些警戒之语来得太迟了，因为这些从哇尔采来的*PM*串已经被顽固迷信的库德尔人吞食殆尽了。

没过多久，远近便回响起令人惊恐的呻吟声；敏感的阿尔弗伯特挡住了自己的眼睛，不愿意目睹这起惨剧。当他最终鼓起勇气张开眼睛时，它看到了一幅悲伤的画面；在它的四周，就它那只唯一的眼睛可见之处，库德尔人毫无生气的躯壳四散躺倒在地，而就在刚刚，他们还在傻头傻脑地狂欢畅饮呢。“他们要是听我的话，何至如此！”善良的阿尔弗伯特悲哀地心想，满是困惑地挠着自己的大白头。它边想边迈着沉重的脚步走回了它那艘外观奇怪的停在北尔极的橘色飞船，最后看了一眼奥光星球上库德尔人尸横遍野的荒凉景象，最后按下了飞行器人造皮仪表盘上的“起飞”圆按钮，飞向了下一个未知的目的地。

当与宴者开启他们仪式性的狂欢时，阿尔弗伯特因恐惧而晕了过去；直到此时，它才清醒过来。它首先听到了身边四处环绕的兴奋的喊叫，然后，当它鼓起勇气张开眼睛时，它看到了一个令人震惊的景象：在它的四周，就它那只唯一的眼睛可见之处，一拨一拨的库德尔人带着清楚无疑的欢愉神色，盯着在它的大白头上方某处的某个正在移动的东西。它抬起头，想看看那到底是什么，只来得及在极短的一瞬间，瞥见一个纤细的形状，夹着一阵奇怪的、尖锐的沙沙声，急速地向下坠落

简要说明

我在这里要向安布罗斯·比尔斯（Ambrose Bierce）致歉，因为我对他的短篇小说杰作《枭河桥事件》（*An Occurrence at Owl Creek Bridge*）中的精彩情节进行了一次粗线条的模仿。但是我的意图是好的。我这则相当简单的寓言，招募了阿尔弗雷德·怀特海和伯特兰·罗素（联合化名为阿尔弗伯特）与库尔特·哥德尔（化名为库德尔）作为主演。我的这则寓言正是要完全颠覆这出经典的悲喜剧，而我的做法是设置了一种奇怪的生灵，让他们没有能力想象*PM*串中具有任何数论意义的观念，可是又能够把这些串视为有意义的信息——也就是说，他们只能看到高层级的哥德尔式的意义。这大概与一个人想当然的预见截然相反，因为*PM*符号系统的发明，原本明明白白就是为了记录有关数字及其特性的陈述，而显然并<u>不是</u>为了写出有关于其自身的哥德尔式的陈述！

为了避免这则寓言有可能产生的其他混淆，我在这里有必要多加几句评述。首先，任何一个言说其自身特性的*PM*串（哥德尔的酷哥当然是一个典型），其长度都不仅仅是我在寓言开头所说的“极长”而已；它的长度是难以构想的。我从来都没有费心计算过，如果用纯粹的*PM*符号写出来，哥德尔的串会包括多少符号，因为我甚至都不太知道这次计算该如何入手。我猜，它包含的符号数很可能会超过“葛立恒数”（Graham’s Constant）(注)，后者通常被引用为“数学证明中出现的最大数”；而即便它不如葛立恒数那么大，也势必可以与之一较短长。所以，要说让一个人直接去阅读生长在奥光星球上的那些串，不管是在低层级上阅读关于数字的陈述，还是在高层级上阅读关于它们自身可食用性的陈述，都是无稽之谈。（当然，数学符号串在一个遥远星球的丛林

中生长并可以被吃掉，这个想法本身也是无稽之谈，可这就是寓言的特权。）

哥德尔创造他的酷哥陈述，经历了46个逐渐升级的系列阶段；他在这一过程中展示出，原则上，关于数字的特定概念是可以用*PM*记号写下来的。“数字*n*的质因数分解中第*k*个质因数的指数”就是一个典型的此类概念。这个概念依赖于在更早阶段定义过的先在概念，如“指数”、“质数”、“第*k*个质因数”和“质因数分解”（没有一个是*PM*的“内置概念”）。哥德尔从未直白地写出这些概念的*PM*表达，因为要想这么做的话，就得写下一条长到令人发指的*PM*符号链。相反，他给每个单独的概念取了一个名字，即用一种缩写来表达，并在接下来的步骤中加以使用；从理论上说，当有需要的时候，这些缩写都可以拓展成纯粹的*PM*记号。哥德尔一次又一次地利用已经定义的缩写来定义更多的缩写，由此小心翼翼地搭建起一座复杂性和抽象性不断攀升的高塔，步步逼向塔尖，直抵质雅数的概念。

梵语中的肥皂剧

这听起来可能有点深奥和缥缈，所以让我再来打个比方。请想象自己面临着一项挑战，要使用传自古印度的梵语，为“肥皂剧文摘展示架”这一当代术语的意义写出一份清晰的解释。最关键的制约条件是，你必须使用纯粹的梵语，依据它最繁荣时候的样子，而且绝对不可在这门语言中引入哪怕一个新词。

为了详细地表达“肥皂剧文摘展示架”的意义，你不得不首先解释的概念，是电和电磁流、电视摄像头、信号发射器和电视机、电视节目与电视广告，说明洗衣机以及不同清洁公司之间激烈竞争的概念，然后阐明陈腐老套的电视剧每日剧集被传播到千家万户的观念，描绘出观众沉迷于无休无止来回重复的情节而无法自拔的景象，还要介绍食品杂货

店、收款台、杂志、展示架诸般概念……“肥皂剧”、“文摘”和“展示架”这三个词，每一个都可以扩展成一条比自己长出数千倍的由古梵语单词构成的链条。为了表达出这个在现代人看来相当平常乏味的词组，你最终得出的文本将铺满上百页纸。

同理，哥德尔的酷哥串，我们通常习惯用一种超浓缩的形式进行表达，只需要“我在*PM*中是不可证的”这种短语；而如果要用纯粹的*PM*记号把它写出来，它将长得可怕——可不管它的长度多么令人望而生畏，我们依然能够精确地理解它的含义。这怎么可能呢？根由在于它的可浓缩性。酷哥不是一个随机的*PM*符号序列，而是一个具有丰富结构性的公式。正如构成心脏的上百亿个细胞通过相当有序的组织形式而可以被总结为一个词——“泵”，构成酷哥的不计其数的符号，也可以用寥寥几个甄选过的汉字总结出来。

让我们再回到那项梵语的挑战任务，试想我改变了规则，允许你定义新的梵语词汇，并可以用它们进而定义更多的梵语新词。如此一来，“电流”便可以获得定义，并在关于电视摄像头、电视机和洗衣机的描述中使用，而且“电视节目”也可以用来定义“肥皂剧”，等等。如果缩写词可以像这样无限制地叠加在缩写词之上，那么你要为“肥皂剧文摘展示架”做一条梵语的说明，就很有可能不再需一本书那么长的解释了，而只需要几页纸，甚至更少。当然，这么做的话，你就彻底地改变了梵语这门语言，把它向前推进了几千年；可这不就是语言向来发展的方式吗？而且这也正是人类思维运作的方式——把旧的观念合成到新的结构里形成新的观念，而这个新的观念本身又可以在后续的合成中使用，如此往复不绝，无穷无尽，距离基础的原始意象越来越遥远；而那种原始意象，正是每种语言生长于其中的土壤。

收束说明

在我的寓言里，库德尔人和阿尔弗伯特都被假设有能力阅读纯粹的*PM*串——不包含任何缩写的串。既然这些串在某个层级上（库德尔人观察到的层级）谈论了自身，那么它们便同哥德尔的酷哥没什么两样，也就意味着，这些串是无限巨大的——如此描述，是因为找不到更好的说法了（不管怎么说，实用就好）。而这还意味着，任何把它们作为关于数字的陈述而进行解读的尝试，都永远不可能给出任何可以理解的东西。因此，正如寓言所描述的那样，阿尔弗伯特的能力是完全不可能的。但是话说回来，库德尔人的能力也是不可能的，因为他们同样也淹没在一片无边无际的符号海洋之中。不管是阿尔弗伯特还是库德尔人，他们唯一的希望，就是注意到某些特定的模式在那片符号海洋中得到了反反复复的使用，再给这些模式命名，从而把串压缩成某种更容易掌控的东西，然后在一个新的、更短的层级上，再次执行这种寻找模式并压缩的过程，每一次都压缩得更短，把压缩不断地推进，直到整个串最终削减成一个简单的观念："我是不可食用的"（或者把它从寓言中翻译出来，就是"我是不可证的"）。

伯特兰·罗素在思考*PM*串时，从未想象过这种类型的层级转换。他受困于一个可理解的前概念，即有关自然数的陈述，不管其有多长或多复杂，总会保持一种标准数论陈述的味道，比如"质数的个数是无限多的"或者"在斐波那契数列里只存在三个完全平方数"。他完全不曾料到，有些陈述竟然具有如此错综复杂的等级结构，使得它们表达的数论观念感觉上不再像是关于数字的观念了。根据我在第11章中做出的论述，一条狗根本无法想象或理解，特定的大型的色点矩阵可以经过结构化的组织，便会变成人、房子、狗和很多其他事物的图像，而不再仅仅是巨大的色点集合了。更高的层级在知觉上凌驾于更低的层级之上，并在这一过程中变成了二者中"更加现实"的一方。更低的层级则被遗忘并埋葬在混沌之中。

这种向上的层级跃迁是一种深刻的知觉转变，当发生在一个不熟悉的、抽象的背景下时，如在《数学原理》的串的世界里，它听起来会非

常不可思议；可是，当发生在一个熟悉的背景中（比如一个电视屏幕上）时，它却显而易见，稀松平常。

我写下这则寓言，是为了演示一次看起来十分不可思议的向下的等级转换。库德尔人在特定的庞大*PM*串中只能看到如“我是可以食用的”这样的高层级意义，而且据说无法想象在那些串中还同时存在任何低层级的意义。在我们这些了解《数学原理》中符号串缘起初衷的人看来，这听起来像是一种不可理喻的僵化偏见；可是，当我们理解我们自己的本性时，却发生了180度的大转弯，因为与库德尔人十分相似的偏向高层级（且仅止于高层级）的僵化偏见竟能风行弥漫，甚至定义了“人的境遇”。

受困于高层级

向下、向下、再向下，下移到我们大脑的神经层面，同时放慢、放慢、再放慢，慢到我们可以看见（或者至少可以想象）在每一个突触间隙中喷射的每一种化学物质——这种似乎可以瞬间清空大脑活动中一切符号品质的巨大知觉转换，对于我们这些有意识的、自觉的、由“我”驱动的人类而言，几乎是无法想象的。那里不再有任何意义，也没剩下一点儿黏糊糊的语义汁液——只留下天文数字那么多的无意义、无生命的分子，终日毫无生气且无意义地向外喷射。

如果你拥有的是一个典型的人类大脑，幸福满满地对其微小的物理成分和它们晦涩难解的数学化微观运转模式一无所知，而是活跃在距其无限遥远的层面之上，与“肥皂剧”、“春季大甩卖”、“超级时装秀”、“SUV汽车”、“SAT考试”、“脏话”、“圣诞老人”、“水上乐园水肺特别演出”、“潜水镜”、“雪球”和“性丑闻”（别忘了还有“人渣”）打交道，那么你的大脑就能够给自己的本性编出一个尽可能令人信服的故事，这个故事的主人公并不是大脑皮层、海马体、杏仁体、小脑或者其

他任何拥有奇怪命名的胶黏性物理结构，而是由一个在解剖中不可见的含糊暧昧的“我”所扮演，辅助并支持“我”表演的其他角色演员是我们所知的“观念”“思想”“记忆”“信念”“希望”“恐惧”“意图”“欲望”“爱”“恨”“对抗”“嫉妒”“同情”“诚实”等等——在这些演员所在的柔软、缥缈而与神经学无关的世界里，你的典型的人类大脑把它自己的“我”视为推动者和原动力，从未有一刻怀疑过，它的明星主演可能仅仅是一个速记法而已，代表的是海量的微小实体以及发生在它们之间的不可见的化学反应，每秒以数十亿计——不对，数千万亿。

于是，人类的境遇与库德尔人的境况便具有了深刻的类比性：两个种族都看不到也无法想象一个更低层级的现实，尽管其对于他们的存在而言至关重要。

怪圈的第一种关键成分

在一个视频反馈系统中，不管出现在屏幕上的形状如何回旋、多么繁密、嵌套得有多深，都不可能发展出一个“我”来。这是为什么呢？答案很简单：一个视频系统，不管其具有多少像素或色彩，都绝不可能生成任何符号，因为视频系统不会感知。在一个视频反馈环的回路中，永远不会有任何可被触发的符号——没有概念、没有范畴、没有意义，并没有比一个音频反馈环刺耳的声音强出一点儿。一个视频反馈系统不会为它屏幕上呈展的奇怪星系形状赋予任何能导致任何事发生的因果力。实际上，它没有为任何事物赋予任何东西，因为一个完全不涉及符号的视频系统，不可能、也永远不会思考任何事物！

由此可知，在视频反馈系统中无法出现的怪圈之所以会在大脑中出现，是因其具有一种能力——思考的能力。实际上，这大可以归结为一个可以表示其拥有一套足够大的可触发符号集合的单字词。自然数的丰富性赐予*PM*表征具有无限复杂性的现象之力，因而得以经由哥德尔的

建构（construction）而扭转回头并吞食自身；同样，我们可扩展的符号集合也赋予我们的大脑表征具有无限复杂性的现象之力，因而得以经由一个怪圈而扭转回头并吞食自身。

怪圈的第二种关键成分

但是，以上所有一切还存在一个反面，那正是让人类大脑中的环路模式称得上“怪”，并让一个“我”从天而降的第二种关键成分。颇具讽刺意味的是，这个反面恰恰是一种无能——是我们的库德尔人无法窥见低于符号层面之物的那种无能。我们都无法以任何方式看到、感知到或察觉到构成我们思维底部的一切感觉不到的冒泡与沸腾；那些微观物质激烈狂热的蒸腾与翻滚，也属于这种无能。我们对于微观世界天生的视而不见，迫使我们产生了一种幻觉，幻想出一道深不见底的鸿沟，隔开的一边是由球、棍、音、光组成的没有目标的物质世界，而另一边则是由希望、信念、欢乐和恐惧构成的抽象世界，这个抽象世界充斥着目标，似乎处于一种从根本上种类不同的因果性的支配。

当我们这些拥有符号的人类观看一个视频反馈系统时，我们自然而然会把注意力放在屏幕上吸引眼球的形状之上，并忍不住想给它们贴上诸如“螺旋走廊”或“星系”等酷炫的标签。可我们心里终究还是很清楚，它们只是由像素构成的而已，不管出现在我们眼前的图案是什么样，它们都仅仅遵循着像素的局部逻辑。这种简单而清晰的认识，剔除了那些酷炫的分形格式塔图案具有任何显在生命或自主性的可能。我们不会费心为屏幕上的涡旋形状赋予欲望或希望，意识就更不用提了——这就如同我们不会费心把天空中漂浮的那些毛绒棉球认作某位艺术家剪影的展示或者某位烈士受刑的情景。

尽管如此，当我们观察自身时，情况就大不一样了。当我们谈论我们自己时，情况比我们谈论视频反馈时要暧昧许多，因为我们在我们的

大脑中找不到任何可以类比于像素及其局部逻辑之物。我们的知识可以告诉我们，大脑是稠密的神经元网络，可是这并不能让我们更熟悉那一层面上的大脑。这就好比我们知道法语是由罗马字母写出来的，可是这并不能让我们成为法语诗歌方面的专家。我们这种生物，生来就没有能力把眼神聚集在推动我们的思维运行的微观机制之上——遗憾的是，我们还不能溜达到街角的一家药店里，买一副便宜的眼镜戴上，来矫正这种视力上的缺陷。

有人可能会质疑说，神经科学家跟普通的门外汉不一样，他们对于大脑的低层级硬件那么熟悉，应该已经懂得如何思考如意识和自由意志等谜题了吧。可通常的情况与此恰恰相反：很多神经科学家正是因为对于大脑的低层级方面太过熟悉，反而更拿不准意识和自由意志到底能不能在物理的层面上得到解释。心灵与物质之间不可逾越的裂断对他们造成了深深的打击；困惑之中，他们完全放弃了发现意识与自我如何在物理过程中发生的努力，纷纷缴械投降，皈依成为二元论者。看到科学家以这种方式投子弃局，的确引人唏嘘，可是这种情况却太常见了。这个事情告诉我们的核心道理是，成为一名专业的神经科学家，绝不等同于对大脑拥有深刻的理解——就好像成为一名职业物理学家也不代表着就能对飓风有深刻的理解一样。事实上，有时，深陷于细节知识的泥淖之中，正是阻碍达到深刻理解的关键所在。

我们人类天生无能，无法窥见在我们大脑内部某个特定层级之下的东西，这就让我们的内部具有了一种可类比于电视屏幕上旋转星系之物——“我”的庞大涡旋星系——其之于我们，乃是一种无可否认的<u>因果性之所在</u>（locus of causality），而非仅仅是由更低层级而来的被动的副现象（如一个视频回路中的星系图案）。我们在头脑中感知到的那个“玻璃球”的坚硬球形给我们造成的印象过于根深蒂固，以至于我们把它认可为与一切我们所知事物同样伟大的真实之物。而且，因为“我”的符号的锁定不可避免地在人类自我感知的反馈环中年复一年地发生，所以这一因果性又调转过来，把“我”请上了驾驶员的座位。

总之，这两种成分的结合——一种能力和一种无能——引发了自我的怪圈；这是一个我们所有人类都会落入的陷阱，无一例外，不问情由。虽然它在刚开始的时候跟低调的马桶浮球机械装置或视频反馈环一样天真无害，没有引入任何反直觉的因果性类型，可是人类的自我感知终将不可避免地呈现出一个对世界施以自上而下因果性的实体把这一信念不断地巩固加强，并完成对它终极的、不可动摇的、不可改变的锁定。最后的结果，常常是对于一切其他视角可能性的全盘猛烈否定。

斯佩里，反诸自身

我刚刚说，我们都落入了一个“陷阱”之中，但其实我的看法并没有这么悲观。保守地说，这样一种“陷阱”并不是有害的；进一步说，我们应该为它感到欢喜，并且对它加以珍惜，因为正是它让我们得以为人。请允许我再引用一次罗杰·斯佩里的动人言论：

> 在这里提出的大脑模型中，一个观念或一种理想的因果潜力变得同一个分子、一个细胞或一次神经冲动一样真实。观念引发观念，并协助新的观念发生演变。它们彼此之间交互作用，并与其他思维力量交互作用。这些思维力量可能来自同一个大脑内部，也可能来自邻人的大脑，而得益于全球化的交流，还可能来自距离遥远的外国人的大脑。此外，它们还与外部的环境发生交互作用，在进化的过程中形成突破性进展，远远超过任何先在的进化景观，甚至超过了有生细胞的出现。

说到底，斯佩里在这里所做的一切，就是冒着风险，在一次严肃的科学发表中斗胆宣明了大街上随便哪个人都认同的一种无聊乏味、稀松平常的共识信念：被我们称为“我”的那个东西，是一个货真价实的真实存在。在科学的世界里做出这样一种声明是要承担遭受质疑围攻的危险

的，因为它听起来太过肤浅，似乎隐隐透出了笛卡儿二元论的腐味（当我在读他这段话时，心里时不时地会蹦出如“生命冲动”（elan vital）、“生命力”、“蜂巢精神”（spirit of the hive）、“生命的原理”（entelechy）和“整体子”（holon）这些美妙而充满神秘色彩的词汇）。

不管怎么说，罗杰·斯佩里心里十分清楚，他并不支持二元论或者任何形式的神秘主义，因此他才有勇气做出这一声明。他的立场维持着一种微妙的平衡，而我确信，他的洞见总有一天会受到世人的认可与歌颂，被类比于库尔特·哥德尔的微妙平衡；而后者阐明的是，在一个形式数学系统中，高层级、呈展式、自我指涉的意义如何能够具有一种与系统严苛、固化、低层级的推理原则同等现实的因果潜力。

---

1. I suppose almost any book on the brain will convince one of this, but [Penfield and Roberts] did it to me as a teen-ager.
2. “right stuff”，这是约翰·塞尔在形容人脑与电脑的差别时挂在嘴边的一个口号，他认为人脑是由“正确的材料”构成的。——译者注
3. 治疗类选法（triage），是一种通常用于战争和灾难后的医疗急救方案，根据需要接受急救的患者受伤严重程度迅速确定优先治疗的等级，以确保最大的存活率。其中隐含的一层意思是，有可能为了提高整体存活率而牺牲某些重伤病人接受治疗的机会。文中以此来类比大脑对于环境信息的分级和筛选。——译者注
4. For some of Moravec’s more provocative speculations about the near-term future of humanity, see [Moravec].
5. 约翰·塞尔谈到了“正确的材料”，支持他所谓的“大脑的语义因果力”，塞尔用这个好听而暧昧的术语来表示，当一个人类的大脑让它的主人说话的时候，那些话不仅仅看起来代表某种东西，而是真的代表某种东西。比如，诗人迪伦·托马斯的大脑。遗憾的是，说道迪伦·托马斯，他说出的绝大多数的话虽然更加动听，但是太过于暧昧不清了，以至于让人纳闷到底是什么样的“材料”才能构成那些话背后的大脑。
6. See [Wells 1986].

第15章
缠结

大脑中的多重怪圈

两章以前，我曾宣称，在每个人类的大脑中，都有一个怪圈，那个怪圈构成了我们的“我”，可我同时还提到，那只不过是粗糙的第一步假设。确实，那个说法实在是太过简化了。既然我们在自己的大脑中都在以水平迥异的细节和忠实度来感知并表征着数百个其他的人类，既然所有那些人类最重要的面向就是他们<u>自己</u>对于自身的认识，那么我也无可避免地要在我们的头脑中反映并容纳数量巨大的其他怪圈。可是，当我们说，每一个人的头脑都是多重的“我”时，这到底意味着什么呢？

好吧，我也说不准那到底意味着什么。我倒是希望我能明白！而且我想，要是我真能明白的话，我就能集世界上最伟大的哲学家和心理学家于一身了。我站在与这座圣山峰顶相去甚远的下方能做出的最好猜测就是，我们制造出一个极简版的属于我们<u>自己</u>的怪圈，并把它安装在我们用来表示他者的符号的中央，让那个最开始相当粗陋的环路结构随着时间而变化和生长。当涉及我们最熟悉的人时——我们的配偶、父母、兄弟姐妹、子女和密友，每一个怪圈，搭配着成千上万独特成分的装饰，都会在岁月的推移中成为一个非常丰富的结构，而且每一种结构都会获得一种极大的自主性，脱离于作为其种子的简化版“原装”怪圈。

没有内容的反馈环

回顾我们之前讲过的音频反馈环隐喻，可以更好地帮我们理解这一“原装”怪圈的观念。设想一个麦克风和一个扩音器被连接了起来，导致哪怕只是一点十分轻柔的噪声，也能快速地飞转起来，每绕过一圈就更响一分，直到变成一阵音量巨大的刺耳尖叫。但是再设想，那间房间从刚开始便是一片死寂，鸦雀无声。如果是这样的话，会发生什么情况呢？会发生的是，它将保持一片死寂，仍旧无声。这个环路的运转一切良好，可是它不会发出任何噪声，因为0乘以任意数字还等于0。当没有信号输入一个反馈环时，这个环路就不会产生可感知的效果；它的存在与不存在甚至都没什么大不同。一个单独的音频环路，不会发出刺耳声。要想让事情开动起来，就必须得有非空的输入。

下面，让我们把这个场景挪用到视频反馈的世界中去。如果一个人把电视摄像头指向一块空白屏幕的中央，而且摄像头收录的只有屏幕，而不包括它的边框，那么不论摄像头是竖立不动、倾斜、转动还是放大、缩小（只要保证不涉及屏幕的边缘），这套装置即便能够形成环路，也只能产生一幅固定的白色画面。同前面的情况一样，来自一个封闭反馈环的图像结果也没有任何变化，因为没有任何来自外部的东西为环路提供内容。我打算把这种没有内容的反馈环命名为“原装”环路。显然，在两个原装视频反馈环之间是无法做出区分的——它们都没有可识别的特质，没有“个体性”。

不过，如果摄像头向左或向右移动得足够远，或者把镜头缩得足够短，可以收入某些外在于空白屏幕的东西（哪怕只是一小撮颜色），屏幕上就会有一点不再是空白的了，这样一来，那一小块并非空白的图像就会立即被吸入视频反馈环中，如同一根被飑风卷走的枝条，一圈一圈地环绕起来。很快，屏幕上就会挤满了色点，形成一种复杂而自成稳定的图案。单凭图像包含其自身这一事实，并不足以为这个原装环路赋予可识别的特征，还有同样至关重要的另一个事实是，图像中还得包含经过特定排列的外部要素。

如果我们把这个隐喻放在人的特质的语境中，我们可以说，一个“光秃秃”的怪圈无法生成一个清晰可辨的自我——它只是一般化的、原装的空壳，要与世界中的其他事物发生接触，才能开始获得一种与众不同的特质、一个独特的“我”。（如果你也属于那些享受非良基集合——忤逆罗素的主张而包含自身的集合——带来的禁忌快感之人㊟，请看我下面这道谜题：两个单元素集合*x*和*y*，每个集合都包含且仅包含其自身，那么*x*和*y*到底是相同的实体，还是不同的实体？回答这道题时，如果规定，当且仅当两个集合拥有相同的元素时，它们才是相同的，那么人们立刻便会陷入一种无限回归当中，无从作答。而我更愿意快刀斩乱麻，百无禁忌地宣布，既然两个集合彼此之间难以区分，所以它们就是相同的集合。）

婴儿反馈环和婴儿的“我”

虽然我刚刚造出了一个人类大脑中的“原装”怪圈的概念，可是我的意思决不是想说，人类婴儿在出生之际便已经被赋予了这种“光秃秃”的自我怪圈——这是一个纯粹的、蒸馏过的“我”的壳，虽然是原装的，却已经得到了完全的实现，而这种天赋单纯是得益于它所拥有的人类基因。甚而说，一个未出生的人类胚胎，尚在腹中时（更别说在受精阶段了！）便已经获得了一个光秃秃的自我环路，那更非我意。人类自我的实现，绝不是如前面说的那样自动化完成且由基因先天决定的。根本不是。

人类自我怪圈的闭合深深地依赖于一种等级变换的跳跃，亦即感知，感知也就意味着范畴化。因此，一个有机体的范畴化装备越丰富、越有力㊟，它的自我也就越完满、越丰富。反之，一个有机体的范畴集合越匮乏，其自我也就越贫瘠，直到抵达极限处，自我便完全不在了。

我已经强调过很多次，蚊子从根本上便不具备符号，因而也就压根儿没有自我。在一只蚊子的脑袋内部，没有怪圈。蚊子的情况也同样适用于人类婴儿，人类胚胎更是如此。只不过，人类的婴儿和胚胎因为享有人类的基因，于是具备了一种神奇的潜力，有可能成为在未来几十年中逐渐扩大的巨型符号库的居所，而蚊子则没有这种潜力。蚊子生来便缺乏符号库，也肯定不会发展出一个符号库，因而它们注定是没有灵魂的（哦，好吧——可能有0.00000001亨内克的意识，比温度计高出的水平几乎可忽略不计）。

无论好坏，总之我们人类在出生的时候，关于我们的感知系统在我们与世界长达几十年的交流中将会变化成什么样子几乎毫无头绪。降生之初，我们的范畴库是如此单薄，出于实用性的考虑，我大可以说它根本没有。因为没有符号可以触发，所以在威廉·詹姆斯（William James）口中栩栩如生的那种感官输入的“巨大的、盛开的、嗡鸣的混沌”[注]，是婴儿无法明白的。对于一个婴儿来说，自我符号的建构尚是一件很遥远的事；所以，在婴儿内部没有自我的怪圈，或者近乎没有。

直白地说，一个人类的新生儿不管有多么可爱煞人，因为它未来的符号机制99%是缺席的，所以它就是没有“我”——或者说的客气一点，就算它拥有最少量的“我”，也只有1亨内克左右那么多，并没什么值得大书特书之处。于是，我们亲眼见证了一个人类的头脑可以包括<u>少于</u>一个的怪圈。那么，要是<u>多于</u>一个呢？

缠结的反馈环

为了用具体的方式探讨两个怪圈在一个头脑中共存这个观念，让我们先从那个老掉牙的电视隐喻的一个轻微变体出发。假设有两个摄像头和两台电视机被架了起来，摄像头A为屏幕A馈送信号，而距离它很远的摄像头B则为屏幕B馈送信号。不仅如此，摄像头A还收入了屏幕A上

的所有图像（外加某些临近的东西，给A的环路提供了“内容”）并将其返送给A；与之类似，屏幕B也收入了屏幕B上的所有图像（外加某些外部的内容）并将其返送给B。如此，既然依据我们的设定，A和B两个系统是远远分开的，那么显而易见的是，A和B构成了两个相互独立的反馈环。如果摄像头A和B拍摄到的局部场景不同，那么屏幕A和B就将呈现出清晰可辨的不同图案，所以两个系统的“特质”也很容易就能够区分清楚。到此为止，这个隐喻带给我们的依然是陈酒（事实上，是两瓶陈酒）——两个不同的头，各自内部有一个环路。

可是，当系统A和B逐渐拉近距离，开始彼此交叠之后，又将发生什么呢？摄像头A看到的将不仅仅是屏幕A，还看到了屏幕B，因而环路B也会进入到环路A的内容当中（反之亦然）。

让我们根据看起来比较自然的情况假设，摄像头A离屏幕A的距离比离屏幕B的距离更近。那么环路A在屏幕A上占据的空间就会比环路B更大，也就意味着环路A占据了更多的像素，所以环路A在屏幕A上的再现，还原度也会更高。环路A将是大而精细，而环路B则是小而粗糙。但这只是屏幕A上的情况。在屏幕B上，一切都颠倒过来：环路B将是大而清晰的，而环路A则是小而粗糙的。在我们继续下面新的内容之前，我想要提醒你的最后一件事是，现在的环路A虽然还是仅仅被叫作“A”，却不免也包含了环路B在内；这两个环路各自在另一个环路的定义中发挥了作用，只不过环路A在它自己的定义中发挥的作用比环路B更大而已（反之亦然）。

我们下面把隐喻用在两个人的身上。A和B两人各自拥有人格同一性（即他们自己私有的怪圈）——而那种私有同一性中有一部分来自另一个体的私有同一性，因而也依赖于对方。不但如此，每个屏幕的图像在另一个屏幕上呈现得越忠实，两个环路的“私有”同一性就纠缠得越厉害，而他们也就更容易开始融合，变得模糊不清，我们甚至可以生造一个词，说他们越来越无法解缠。

到了这一步，我相信，即便我们是单纯沿着一个十分奇妙的科技隐喻前进，我们距离理解真正的人类同一性是怎么回事也在不知不觉中慢慢地靠近了。事实上，有谁可以想象，在始终未遭遇某种陌生的抽象结构的前提下，也有可能获得穿透人类同一性之谜的深刻洞见呢？西格蒙德·弗洛伊德设立了本我、自我与超我，而在一个人类灵魂的构造内部，完全可能存在这样的抽象物（也许并不一定是这三种，但也差不多是那种模式）。我们人类与其他的自然现象竟如此不同，甚至与大多数其他物种的生灵也相差迥异：我们竟然预想到，为了对于我们自身的真相有所了解，我们必得向完全意料之外的地方去寻找答案。虽然我的怪圈与弗洛伊德的主张显然不尽相同，但在精神上却有某种共通之处。两种关于自我为何物的观点都包括了离自我所居的生理基底相去甚远的抽象模式——事实上，这距离远到让这基底的特性看起来已经无关紧要了。

我们头骨内部一个处于优势地位的环路

设想某种未来的电视技术做到了可以抹除摄像头和屏幕上的粗糙质地，令所有的图像都能完美无瑕地再现出来。在这种神奇的设定下，我们上述的论点将不再成立，即A对于环路B的呈现，因为使用了更少的像素，所以不如它对自己的环路的呈现那么忠实。现在，A在它的屏幕上完美地呈现了环路B，反之亦然。这样一来，A和B的不同之处在哪里？或许它们现在已经不可区分了？

话不能这么说。即便A和B彼此都能完美地呈现对方，二者之间还是有根本性的差别。这个差别就是，摄像头A把它的成像直接馈入屏幕A（而非屏幕B），而摄像头B则把它的成像直接馈入屏幕B（而非屏幕A）。如此一来，如果摄像头A倾斜或拉远，那么屏幕A上的图像也会随之更改，发生倾斜或扩大，而屏幕B上的图像则会保持原样。（当然，屏幕B上内嵌的屏幕A图像也会随之倾斜或扩大，一直影响到嵌套最深

的图像——但是系统B中顶层屏幕的方位和大小却是保持不变的，不会像系统A中的顶层屏幕上的图像一样，直接受到摄像头A的运动的影响。）

考虑这种变化形式的要点是为了澄清，即使在一个环路与环路深度缠结的场景中，可以明确区分的特质依然存在，因为一个系统的感知硬件只为那个系统提供直接的馈送。它也许会对所有各式各样的其他系统发生十分重要的间接的影响，但是任何感知硬件都首先且主要与它直接馈入的系统（或者用当今计算机科学与神经科学含糊其词的行话来说，是与它“硬连接”的系统）联系在一起。

直白一点说，我的感知器官直接向我的大脑进行馈送。它们同时也馈入我的孩子、朋友和其他人的大脑（比如，我的读者），但这都是间接进行的——通常是经由中介性的语言渠道（但有时候也经由照片、美术或音乐等形式）。我给我的孩子讲述发生在食品杂货店收银台的某件古怪趣事，而确如其实，这件事马上就会在他们的脑海之中一清二楚地浮现出来！一位顾客的购物车里有一期黑白的《世界新闻周刊》（*The Weekly World News*），收银员把报纸拿出来，当她看到头条新闻上写到那个被救的婴儿毫发无损地漂浮在一艘来自泰坦尼克号的救生筏里时，脸上露出了古怪的表情，那位顾客尴尬地轻轻笑了笑，排在后面的顾客说了一句俏皮话等画面在我的孩子、朋友和他人的大脑中被勾勒了出来，有时候似乎如直接通过他们自己的感知器官收到的画面一样栩栩如生，毫不逊色。

我们以这种间接的方式感受生活的方式真的是人类交流中一个无比美妙的方面，但是任何人的知觉输入当然绝大部分还是来自他们自己的感知硬件，只有很小一部分是以这种方式经由他人的经验过滤得来的。直白地说，正因如此，我才得以基本上还维持着我自己，而你也基本上保留着你自己。然而，如果我的感知像它们冲入我的大脑一样迅速而猛烈地涌入你的大脑，那么我所谈论的局面就完全相同了。但至少就眼下

而言，可以说，在我的眼睛和你的大脑之间，还不存在如此之高的交流频率的风险。

共享的感知，共享的控制

我首先提出，一个人类的“我”产生于人类大脑中一个十分特别的怪圈的存在。可是如今我们看到的是，既然我们在自己的大脑中镜照了很多人，也就存在着很多尺寸不同而复杂程度各异的环路，因而我们不得不改进我们的认识。如我刚才所说，这种改进有一部分取决于下面这一关键的事实，即在一个特定脑中的这些环路里，有一个环路是居于优势地位的——由直接馈入那个大脑的感知系统所引发。可是，这并非故事的全部；另外一部分与感知无关，而是关乎一个大脑所控制之物。

我房子里的温度计并不能调节你房子里的温度。类比而言，在我的大脑中做出的决定也不能控制与你的大脑硬连接起来的身体。当你和我打网球的时候，我的大脑只控制我的胳膊！或者至少，乍一看似乎如此。可转念一想，这显然是太过简化了，而此时，事情又要开始变得扑朔迷离了。我对于你的胳膊有部分而间接的控制——毕竟，我把球往哪个方向打过来，你就要往那个方向跑过去，而我的击球也与你如何挥动胳膊关系甚大。所以，通过某种间接的方式，我的大脑能够在一场网球比赛中控制你的肌肉，但这并不是一种非常可靠的方式。同理可知，如果我在一条马路上驾驶时踩下了刹车，那么跟在我身后的人也会纷纷踩下刹车。在我的大脑中发生之事对于那些驾驶员的行为施加了一点儿控制，可它是一种不可靠而且不精确的控制。

刚刚描述的那种类型的外部控制并不能造成对两个人同一性问题的深度混淆。网球和驾驶无法引出有关灵魂的深度阐释，但是当语言也掺和进来之后，事情就变复杂了。正是通过语言，我们大脑得以对其他人类的身体施加程度可观的间接控制——这不仅对于父母和教官而言是一

个再熟悉不过的现象，而且对于广告商、政治“化妆师”以及死皮赖脸地求满足而花样百出的少年而言也绝不陌生。通过语言，其他人的身体可以成为我们自己身体充满灵活弹性的延伸。从这个意义上来说，我的大脑便也与你的身体牵连在一起，同它与我自己身体牵连的方式多少有点相似——只不过要再重申一次，这个联系并不是硬连接。我的大脑通过和你交流而牵连到你的身体，这种渠道与把它连接到我的身体的那些线路相比要慢很多，也间接很多，所以控制也远没那么有效。

例如，我用我自己的手写我的签名，相比于努力通过描述那么多曲线的所有细节而让你把它写出来，要容易太多了。每当我在食品杂货店收银台“签字”时，我都可以顺其自然且无意识地把这些曲线画出来。但是，我们最初的想法——在我的大脑和我自己的身体相连的方式与它和其他人身体相连的方式之间存在根本性的、绝对化的区分——看起来是有些夸大其词了。二者在程度上有差异，这一点显而易见；然而，它们是否在本质上属于不同的类别，这就不清楚了。

我们有关缠结的灵魂的讨论，已经进行到哪一步了？我们已经看到，我可以间接地感知到你的感知，而且我还能间接地控制你的身体。同理，你也能间接地感知到我的感知（你现在就在这么做！），而且也能间接地控制我的身体，至少可以控制一丁点儿。我们还看到，因为交流渠道的速度足够慢，所以这里有两个泾渭分明的独立系统，我们也能毫无困难地为他们赋予不同的名字。我们人类具有泾渭分明的独立身体这一事实（除了母体-胎儿的共生和连体双胞胎以外），使得为每个身体赋予一个不同的名字成为一件绝对自然之事，而且在一个浅显的层面上，为不同身体分配不同名字的动作似乎一劳永逸地解决了所有的问题。“我，泰山；你，珍妮。”我们的命名习俗不仅支持而且极大地促成了一个令人舒适的信念的锁定，即我们——我们的自我——是截然分明的独立实体。“我，泰山；你，珍妮”——尘埃落定。

可是，在把身体确定为同一性之所在这件事上，语言起到的作用还

不止于此。它不仅帮给我们的每个身体都起了一个名字（“泰山”“珍妮”），还为我们提供了人称代词（“我”“你”）。人称代词能够同名字一样有力地巩固那个信念，即不同的灵魂之间存在一种鲜明而透彻的区分，而每一个密不透风的灵魂都只与一个身体相连。下面就让我们来细细打量一番。

双旋世界中的一场旋风旅行

几年以前，我曾构想过一个奇妙的哲学想象世界，在得到你许可的前提下，我将在后面几个小节中陪同你前往那个世界进行一番游览。虽然我在当时并没有给那个地方起名字，但是我想在这里把它命名为“双旋世界”（Twinwirld）。双旋世界的独特之处在于，在那里出生的人，99%都是同卵双胞胎，只有1%的人是单胎出生的，但那里的人不这么叫他们，而是称他们为“单人”（halflings）。在双旋世界里，双胞胎（跟在我们世界里一样，并非毫厘不差地全等，而是拥有相同的基因组）共同成长，形影不离，穿同样的衣服，上同一所学校，选同一门课，合作完成家庭作业，交同一拨朋友，学同一种乐器，最终作为一个小组共同承担一份工作，如此等等。在双旋世界中，一对双胞胎被称为一个“偶人”（pairsons）或“可分人”（甚或干脆是“双体人”）。这也是在所难免。

双旋世界中的每一个可分人在出生时都会被赋予一个名字——于是一个男性的偶人可能被命名为王伟，而一个女性的偶人则可能被叫作孙玥。如果你非想知道的话，要想指称一个偶人的两“半”中的每一半，也是有办法的，虽然这种需求碰巧是微乎其微的。不管怎么说，出于完整性的考虑，我还是会描述一下该如何做。你只需要在可分人名字的后面加上一个连字符，再接上一个三笔画的部首——要么是“工”，要么是“口”——就可以了。（双旋世界的词源学家已经认定，这两个部首“工”和“口”并不是随意选定的，事实上是“左”和“右”这两个词的残

遗，虽然没有任何两人能确定到底为什么应该是这样。）如此而言，王伟就包括了一个“左半”，王伟-工，和一个“右半”，王伟-口。同理，孙玥也由一个孙玥-工和一个孙玥-口构成——但是正如我所说，绝大多数时候，没有人们会感到有需要强调一个偶人的“左半”或“右半”，所以这些后缀几乎从来都用不上。

话接上文，在双旋世界中，一个“朋友”又是由什么构成呢？当然是另一个偶人——你你（“UU”）很喜欢的某两人。那么，爱情和婚姻呢？好吧，如果你已经猜到一个偶人要与另一个偶人相爱并结婚，那你就正中靶心了！事实上，出于惊人的巧合，我刚刚提到的王伟和孙玥正好是一对典型的双旋世界情侣；不仅如此，他们还骄傲地成了两个偶孩的父母——一个名为“晓静”的女偶孩和一个名为“剑英”的男偶孩。（为了应付八卦爱好者，我不得不解释一句，我不知道是孙玥-工还是孙玥-口生下了那一个偶孩，也不知道，暂且这么说吧，每次生育的始作俑者到底是王伟-工还是王伟-口。在双旋世界中，没有任何两人思考过如此私密之事——就好像在我们的世界里，我们也不会在意，导致一个孩子诞生的精子到底是来自父亲左边的睾丸还是右边的睾丸，或者那颗卵子到底是来自母亲左边的卵巢还是右边的卵巢。反正不是左边就是右边——偶受精卵形成了，偶孩出生了，这才是关键。总而言之，请不要围绕这个复杂的话题提出太多问题了。这离我这个幻想的主旨偏离太远了！）

在双旋世界里，有一种心照不宣的显见共识：虽然每个可分人都是由在物理上可以区别和分离的两半构成的，但是这两半之间的联结已经紧密到使其在物理上的区分可以忽略不提了，所以这个世界的基本单元是偶人，而不是它的左半或右半。每两人都是由一个左半和一个右半组成的，这只是一个关于生命存在的日常事实而已，跟每半部分都有两只手、每只手都有五根手指一样，被认为是理所当然。事物是由部分所组成的，这自然没错，可这并不意味着它们没有作为一个整体的完整性！

一个偶人的左半和右半有时会在物理上离开彼此，但是通常都只分开很短的一段时间。例如，偶他（twey）中的一半可能快速地返回食品杂货店去买一件偶他刚刚忘了买的东西，而另一半则回家做晚饭。或者，如果偶他在一个山坡上滑雪，那么偶他可能在从两边绕过一棵偶树的时候短暂地分开。但是在绝大多数时间里，这两半都更愿意寸步不离地守着对方。而且，虽然这两半也会在一起交谈，可是因为绝大多数想法都很容易被对方预料到，所以其实不需要多少言语，甚至在交流复杂观念的时候也不例外。

“你你”是一个词？

我们接下来要讨论的问题有点棘手，是关于双旋世界里的偶人称代词（pairsonal pronouns）。首先，他们为每一个割裂开的一半准备了一个类似于我们熟悉的代词“我”这样的东西，只不过他们是用“单我”（i）来表示。这是因为“单我”和后缀“工”与“口”一样，也是极少用到的说法，只有在极端的学究式的论证中才会用到。远比“单我”更为普遍的代词是偶人中任意一半用来指称整个偶人的代词。我说的不是代词“我们”，因为这个词超出了正在说话的偶人的范畴，还包括了其他的偶人。因此，举例而言，“我们”可能意味着“我们整个学校的全部成员”或者“昨晚派对上的每两人”。与之相反，还有一个“我们”的特殊变体——“偶我”（Twe）——正是用来指称说话的那个左半或右半所属的偶人。当然，还会有一个类似的代词，“你你”（虽然它看起来似乎应该发音成连续的两个“你”，但它实际上却读作“偶你”），用来称呼一个偶人。因此，举例而言，当王伟和孙玥初识彼此时，王伟（不是王伟-口就是王伟-工——我不知道是偶他中的哪一个）曾战战兢兢地对（偶他迷恋上的）孙玥说：“孙玥，今天晚饭之后，偶我想去看场电影，你你想加入偶我吗？”

代词“你”在双旋世界中也是存在的，但它只能用在复数“你们”中；

这也就意味着它永远不会被用来称呼另一个可分人——而总是用来指称一个群体。“你们会滑雪吗？”可能问的是整个家庭，但绝不是在问一个偶孩，或者一个偶父或偶母。（如果要问的话，那当然是：“你你会滑雪吗？”）与之类似的是，“他们”也绝不会仅仅指一个可分人。“他们两个都参加了我的婚礼”说的是一双偶人（也就是一共有四半——或者根据在我们的世界中流行的古怪说法，就是四个“个人”）。至于第三人称单数代词，也有一个——“偶他”——是不分性别的。于是，“偶他昨晚去听演唱会了吗？”就是在询问有关孙玥或王伟（但不是他们两对同时，因为那就需要用“他们”了）的事，而“偶他得麻疹了吗？”问的要么是李博，要么是赵静，反正不是二者同问。

双旋世界中的偶人同一性

在双旋世界中，一个年轻的偶人在成长中会逐渐形成一种仅仅作为一个单元而存在的自然认识，虽然偶他是由两个不相连的部分组成的。“每个可分人都是不可分的”是双旋世界中的一句古老民谚。双旋世界中各式各样的习俗惯例全都在系统地巩固和锁定这种统一而不可分的感觉。例如，你你在学校所做的功课只能拿到一份成绩。也许，你你中的一半比另一半在比如绘画或数学上更弱一点，但这并不影响你你的整体自我印象；重要的是小组的合作表现。当一个偶孩学习演奏一种乐器时，两半都会拥有各自的乐器，并练习同样的曲子，而且步调保持一致。等长大一点之后，在你你的大学时代，你你阅读由偶人们创作的小说，参观偶人们举办的画展，学习由偶人证明的定理。总而言之，不论是嘉奖还是贬斥，荣誉还是耻辱，默默无闻还是一鸣惊人，都总是加诸偶人之上，而绝不会仅仅算在偶人中一半的头上。

两半所组成的偶人是一个自然而不可分解的单元，这个观念构成了双旋世界文化常态的默认前提，因而也得到了进一步的巩固。在我们的社会中，同卵双胞胎常常渴望与彼此分开，闯出自己的一片天地，告诉

世人他们并非一模一样的人；而在双旋世界中，这种欲望或行为则会被视为异数，而且会遭到深深的不解。一个偶人的两半会挠着偶他的头，（或者互相挠着对方的头——有何不可呢？）对彼此说，甚至可能是异口同声地说："偶他到底为什么要分开呢？有谁会想要成为一个单人呢？那种存在简直太不完整了！"

我在一开始就提到过，在双旋世界中，有1%的人生来是单人，而不是偶人。事实上，这个比例并不是精确的1%——更接近0.99%。但不管怎么说，在双旋世界里，一个很年轻的偶人有时会浮想联翩：要是自己生来就是一个单人，并非由两个几乎一模一样的"左半"和"右半"组成，也没有这两半永远缠在一起，说同样的话，想同样的事，形成一个密不可分的小组，那将是一种什么样的感受。偶人的状态看起来实在是太正常了，要去想象一个单人极其奇怪、不完整而贫瘠的生命（通常被戏称为"半条命"），简直太别扭了。

那么，剩余的那微小的一部分、只有0.01%的概率发生的出生情况，又是怎么回事呢？好吧，在怀孕的过程中，可能会发生一种奇妙的现象：构成偶合子的两个受精卵在同时一分为二（没有两人知道它为什么会这样，但它就是会发生），而结果就是，生下来的不是不是单一的偶孩，而是出现了两个基因完全相同的偶孩！（古怪的是，这些婴儿被称为"全等偶双胞胎"，虽然他们从来都不是精确地全等。）偶双胞胎的偶父母（pairenthood）当然会同样喜爱他们的两个"全等"后代，而且还经常会给他们起一些听起来搭配可爱的成对的名字（比如，要是孪生偶女孩的话，就可能是"晓娴"和"晓静"，要是孪生偶男孩的话，就可能是"剑英"和"剑雄"）。

有时候，偶双胞胎在成长的过程中会感受到有与彼此隔离开来的需要，想闯出自己的一片天地，告诉双旋世人，他们并非一模一样的偶人。但是话又说回来，还是有一部分偶双胞胎会乐此不疲地全身心投入到扮演近乎一模一样的偶人的游戏当中。例如，张扬和张帆是一对典型

的偶双胞胎男孩（事实上，他们现在已经长大成人了），他们热衷于让张扬出现在本来应该由张帆出现的场合，或者反过来，搞得朋友们晕头转向。在双旋世界里，几乎每两人都觉得这种小伎俩相当有趣，因为偶双胞胎的经验之于双旋世界中的普通偶人是极为罕见的。确实，双旋世界中的一个正常（非偶双胞胎的）偶人，关于作为偶双胞胎是一种什么样的感受，几乎是完全没有概念的。同一个与俩自我几乎一模一样的某两人肩并肩地成长是一种多么奇怪的事呀！

甚至，有一位双旋世界里的作家曾构想出一个哲学中的幻想世界，叫作“双生双旋世界”（Twinnwirrld），其定义是，在那里出生的人，99%都是“全等偶双胞胎”——可那就完全是另外的故事了。

在双旋世界旋弄下的“偶我”扭旋

我们这场短暂而且希望足够刺激的游览观光，将不可避免地引发一些纠结的问题。其中最鲜明的一个问题当然是，在双旋世界里，一个单人身体——一个单人——构建的是一个作为“单我”的意识，而与此同时，一个偶人身体——一个偶人——也构建起一个作为“偶我”的意识。后一种过程有一部分得益于基因（决定一个偶人的，只是一个在偶合子里发现的基因组），还有一部分得益于文化熏陶，并由大量的语言规范进行了巩固加强；而这种规范，我们在上文也提到过一些。

设想我们打算把蕴含丰富的单词“灵魂”应用在双旋世界里的生命之上。那么在双旋世界里，谁，或者什么，才拥有一个灵魂？甚至连“生命”这个名词也是一个含义丰富的词语。在双旋世界里，一个生命是由什么构成的？在我看来，这两个问题的答案都与下面这个问题的同解：“在双旋世界里，哪种实体构建出了一种把自己认作‘我’的不可动摇的信念？是一个单人，一个偶人，还是二者皆是？”我们真正在问的是，那两个显眼而具有竞争性的类比各自有多么强大——即“单

我”与“我”之间的类比有多么强大，以及“偶我”与“我”之间的类比多么强大？

我猜，凡是阅读本章的人类读者，都可以轻松地对一个双旋世界中的单人（比如孙玥-工或王伟-口）形成认同感，这也就意味着，“单我”/“我”类比在绝大多数读者看来都更有说服力。然而，我希望我的人类读者还能在“偶我”和“我”之间看到一种具有说服力的类比，即便在某些人看来，它并不如“单我”和“我”之间的类比那么强大。不管怎么说，既然双旋世界只是一个幻想，人们大可以根据各自的需要调整它的参数。我和你，我的读者，我们都可以自由地旋弄双旋世界上各种各样不同的旋钮，让“单我”变得更弱，而让“偶我”变得更强，反之亦然。

在我们刚刚经历的那场旋风旅行中，我把决定了双旋世界的旋钮调整到了一个适中的水平之上，让两种类比粗略地达到同等的合理性，因而使得“单我”和“偶我”之间的竞争变得相当激烈。但是现在，我要扭转双旋世界，让“偶我”更强一点。在这个被我称为“连体双旋世界”（Siamese Twinwirld）的新的幻想世界里，我不再设定有99%的人生来都是标准的同卵双胞胎，而是改成有99%的人生来都是连体双胞胎，比如说，都是在臀部相连。除此之外，我还要进一步设定，在连体双旋世界中，双旋世界里的代词“单我”不复存在。这样一来，就只剩下了一种类比，即我们的“我”之概念和他们的“偶我”之概念之间的类比。乍一看，似乎太过荒诞不经了，可奇妙的是，我们标准的现实世界与连体双旋世界竟有大量的共同之处。原因如下。

我们都拥有两个大脑半球（左半和右半），每个半球都差不多可以在另一侧大脑损伤时独当一面，作为一个大脑良好地运转。亲爱的读者，我要假定你的两个脑半球都健康完好，这样的话，当你说“我”的时候，你所指的对象中就包括了一个由左右两个半脑构成的紧密小组，而每个半脑都各自直接接收由你的一只眼睛和一只耳朵传来的馈送。可是，你的小组中两个成员之间的交流迅速而强烈，以至于整个融合的实

体——小组本身——看起来仿佛一物，是一个绝对不可割裂的自我。你对这种感觉再清楚不过了，因为你就这样被建构起来的！而且如果你们跟我属于同类的话，你的两个半球就绝不会四处张扬称自己为“单我”，并且贸然宣布自己为一个自主的灵魂！恰恰相反，它们两个放在一起，才构成了一个“我”。简言之，我们自己在这个世界中面临的人的境遇，同连体双旋世界中的偶人足以类比。

在双旋世界（不管是连体版本还是原始版本）中，一个可分人的两半之间的交流当然不如人脑中两个大脑半球之间的交流那么有效率，因为我们的脑半球是硬连接到一起的。另一方面，双旋世界中单人之间的交流远比我们“正常”世界中任何两个个体之间的交流效率更高。所以，两个双旋世界单人融合的程度虽然不如两个大脑半球融合的程度那么深，可还是比我们的世界里两个关系十分亲密的兄弟姐妹之间的融合更深，比全等双胞胎之间的融合更深，也比夫妻之间的融合更深。

双旋世界，又及

我在写完这一章的初稿之后，就继续开始写下一章。下一章的内容，主要是基于我和丹尼尔·丹尼特在1994年的电子邮件往来。我在他发给我的邮件消息里，发现他曾引用过自己1991年的著作《扩展的意识》（*Consciousness Expanded*）中提到过的一对不同寻常的英国双胞胎（我读过那本书的手稿）。丹的这封邮件我已经忘记了，所以我决定回到他的书里去寻找线索，于是我找到了下面这段文字：

> 我们可以想象……两个或更多的身体分享一个自我。在英国的约克郡可能真的有一个这样的案例：查普林（Chaplin）家的双胞胎，格里塔和弗蕾达（Greta and Freda；《时代周刊》，1981年4月6日）。这对全等双胞胎已经年过四十，共同生活在一家旅馆里，而且作为一个人行动。她们会合作完成说话的行为，例如毫不费力

地补全另一方没说完的句子，或者异口同声说出一样的话，相差不过厘秒。几十年来，她们形影不离、寸步不分，与连体双胞胎无异。

跟她们打过交道的人表示，与她们相处，有一个不言自明的最自然而最有效的策略，就是把她们当成一个她来看待……

我绝不是想说，这对双胞胎具有心灵感应、超感知或者某种超自然的联系。我想表达的是，她们拥有大量微妙的日常交流与合作的方式（事实上，全等双胞胎经常发展出这类精湛的技术）。既然这对双胞胎一生都在看、听、触、嗅并思考大量相同的事件，而且毫无疑问的是，从一开始，她们的大脑就倾向于以相似的方式回应这些刺激，所以她们二人要想维持某种松散的同步，并不需要庞大的交流渠道。（话说回来，就算是我们中间自我控制最强的人，又能保证多大的统一性呢？）……

可是不管怎么说，难道真的不存在两个区别定义的个体自我，各自属于双胞胎中的一方，并协力维护着这种奇妙的双簧把戏吗？可能有吧；但是，要是这两个女人都变得（我们会这么说）如此无私，全身心投入到二人共同的生活当中，并在这个过程中多多少少（我们也会这么说）失去了她的自我呢？

我的双旋世界的幻想在此绽放时已然枝繁叶茂，而我已经记不太清我首次萌生这个幻想的胚芽是什么时候了，虽然我更愿意相信，那是在我读到丹的书里关于查普林双胞胎的段落之前。但是，至于我到底是从丹那里汲取了有关这一观念的灵感，还是完全凭空捏造出来这一幻想，其实都无关紧要；我很高兴不仅丹与这个观念有所共鸣，而且现实人类行为的观察者们还声称目睹过同我纯粹是天马行空的臆想很像的某种东西。双旋世界由此获得了比我料想中更高的合理性。

还有另外一个奇妙现象，在纯属意外的巧合之下，与本章的内容达到了出奇的吻合。在写完双旋世界这一部分之后又过了一两天，我碰巧

在我的床头柜上看到了一小片纸，上面是我的笔迹，用铅笔写下了四个德语词“O du angenehmes Paar”（好生甜蜜的一对佳人啊）。这个简短的短语并没有让我得到什么提示，反而是从那种古调而颂赞的语气中，我猜到它可能是来自巴赫某部康塔塔（cantata）中一段咏叹调的开场词，我在收音机上听到过一次，觉得很合心意，于是就随手记下了。通过网络检索，我很快就证实了自己的猜想——这是康塔塔第197号《上帝是我们的信念》（*Gott ist unsre Zuversicht*）中一段男低音咏叹调的第一句歌词。这原来是一首“婚礼康塔塔”——为结婚仪式的伴奏而作。

以下便是男低音唱给新人的歌词。上面一段是德语原版，下面一段是我的英文翻译，尽量忠实于原文的节奏与韵律：

O du angenehemes Paar,

Dir wird eitel Heil begegnen,

Got wird dich aus Zion segnen

Und dich leiten immerdar,

O du angenehmes Paar!

*

O thou charming bridal pair,

（好生甜蜜的一对佳人啊，）

Providence shall e’ er caress thee,

（神明之光将永远轻抚伊，）

And from Zion God shall bless thee

（而上帝将从天国赐福于伊）

And shall guide thee, e’ er and e’ er,

（并将指引伊，永不迷途，）

O thou charming bridal pair!

（好生甜蜜的一对佳人啊！）

亲爱的读者，你有没有感受到，在这几句歌词里，有某个地方相当怪异？我感受最强烈的地方是，虽然它们是唱给一对新人的歌词，但是却使用了单数代词——德语里的**du**，**dir**和**dich**，㊟和我在英译版本中启用的古代词“thou”和“thee”。在某种层面上，这些第二人称单数代词听上去既不顺耳，用法也不对，可是这些歌词通过用单数称呼一对新人，传达了一种两个灵魂在神圣的结合下瞬间融为一体的深邃感觉。在我看来，这些诗行暗示出，它们所吟诵的那场婚礼形成了一个“灵魂融合器”，促生了只有一个“更高层级灵魂”的单一结构，就像两滴水汇于一处，轻轻一触便天衣无缝地融合在一起，展现出的是一加一有时候仍等于一。

我还找到了这曲咏叹调的法语和意大利语译本，而它们同样也使用了tu（第二人称单数主格）来称呼新婚的二人，而这跟德语一样，听起来比英语版本远为更加古怪，因为tu（在那两种语言中都是如此）完全是现代的标准用法（和“thou”不同），而它总是指称单独一个人的，从来绝不用以指称两个人或者任何种类的一小拨人。

要想在现代英语中体会到同样的语义不协，你就不得不把第二人称换成第一人称，然后再想象一个“we”（我们）在文字使用中的对立面——也就是用“I”（我）来指称他们组成的组合的两个人。接下来，我就要违背事实地把康塔塔第197号延长，想象在婚礼仪式的末尾还有一

句由刚刚结合的两位新人合唱的最后一段欢乐的咏叹调。它的第一句是“Jetzt bin ich ein strahlendes Paar”（如今的我是一对光彩照人的情侣）——而这对新婚夫妇将从头至尾以精准一致的节奏完成合唱，而不是按照典型的巴赫式合唱，每个人演唱不同的旋律；否则就会不合时宜地引人关注他们二人不同的角色。在这段结束曲中，“我”指的是这对夫妻本身，而不是其中任何一方，而且这段咏叹调将被视为这对夫妻的<u>一个</u>新的声音，而不是两个不同的声音。

灵魂伴侣和伴侣灵魂

双旋世界幻想的真正要点在于对我们世界中通常不存在异议的一个刻板信条提出了质疑，用一句口号来表达这个信条就是：“一个身体，一个灵魂。”（如果你不喜欢“灵魂”这个词，那么尽可以选用“我”“个人”“自我”或“意识之所”进行替换。）这个观念虽然很少形诸笔端，却被视为理所当然，以至于在绝大多数人看来，是一句毋庸赘言的废话（除非他们从根本上就拒绝承认灵魂的存在）。但是，参观游览双旋世界（或者如果旅程安排不利，干脆冥想思索也可）却把这一信条推上了风口浪尖，哪怕不被推翻，至少也必须接受对质。所以，如果我能够设法让我的读者放开他们的思维，接受把两个身体联合作为一个灵魂的可能之所这一反直觉的概念——也就是能够像认同《星球大战》中的C-3PO或R2-D2一样轻松地认同一个如王伟或者孙玥这样的偶人，那么双旋世界的使命就达成了。

我的双旋世界幻想的灵感来源之一，正是把一对夫妻视为一种由两个平常个体组成的“更高层级的个体”这一观念；正因如此，我才会说，偶遇写有“O du angenehmes Paar”的那张小纸条是一次惊人的巧合。很多已婚的人都在他们的婚姻旅程中自然而然地习得了这个概念。事实上，我在结婚之前已经凭直觉隐隐约约地感受到了某种类似的道理，而且我至今依然记得，在我的婚礼之前的那充满期待的几周里，我是如何发现

这其实是弗朗辛·柯拉斯本（Francine Klagsbrun）那本《已婚者：在离婚时代不离不舍》（*Married People: Staying Together in the Age of Divorce*）中推动全书的隐含主旨。例如，在讨论已婚夫妇治疗与咨询的一章末尾，柯拉斯本写道："我相信，一名治疗师对待婚姻中的双方，即两位病人，应该保持中立的不偏不倚的态度，但是至于第三位病人，即婚姻，对其表现出偏爱，在伦理上是无可指摘的。"她在为了让婚姻变得更好的治疗中，把婚姻本身也视为一名"病人"，我对于这个观念印象极为深刻，而且我必须承认，在后来的岁月里，这幅图景中的某种真实感，对于我渡过自己婚姻中的艰难时刻助益甚多。

结婚很久的两个人之间形成的那种联系常常无比紧密而有力，甚至其中一方的离世，也能引致另一方过早的死亡。而即便另一方幸存下来，也常常会伴随着一种恐怖的感觉，似乎他们灵魂的一半已经被剥去了。在婚姻的幸福日子里，两个成员当然会有各自的兴趣与风格，但是与此同时，一套共同的兴趣和风格也在慢慢建立起来，而随着时间的流动，一个新的实体也会渐渐成形。

就我的婚姻而言，那个实体就是卡罗尔-道格，时不时地会被戏称为"道卡"或者"卡格"。在我们结婚第一年的某些场合里，我们的合二为一就已经开始在我的思维中清晰地浮现出来了。最早一次是在我们邀请几位朋友来家里参加晚餐派对之后，等每个人都离开之后，我和卡罗尔开始一起打扫房间。我们把盘子收拾到厨房里，然后一起站在水池边上，擦洗、冲刷、晾干，回忆整个晚上发生的事情，直到我们可以在我们联合的脑海中回放，为不自主的小机智会心一笑，反复品味那些意想不到的交谈，评论哪些人看起来很开心，而哪些人看起来有点儿消沉——而在这些后派对的解压缩过程中，最让人惊讶的是，我们两个人几乎在所有观点上都保持一致。有某种东西，某一种东西，由我们两个人共同促生成形了。

我还记得，在结婚几年之后，我们偶然会听到这样一种无比奇怪的

评论："你们长得太像了！"我觉得这简直不可思议，因为卡罗尔在我的眼里美若天仙，至少在相貌上跟我完全没有任何相似之处。可是随着时间的推移，我逐渐开始看出来，在她凝视的眼神中的某种东西，以及关于她如何看待外部世界的方式的某种东西，会让我想到我自己的眼神，以及我自己看待世界的态度。我明白了，我们的朋友在我们身上看到的"相似性"并不在于我们面部的解剖结构当中；相反，它仿佛是某种我们的灵魂向外投射之物，而人们可以观察到的，其实是我们的表情的一种高度抽象的特征。我在我们的某些合影中可以无比清晰地看到这种相似之处。

作为胶子的孩子

然而，我们之间最深刻的联系毫无疑问是由我们两个孩子的诞生所建立起来的。当我们的孩子尚未出生之前，我们还没有完全地融合在一起——事实上，跟大多数夫妻一样，我们时不时地会感到完全不知所措。可是当新的生命走进我们的生命之后，那脆弱的小小的人儿，让我们内部的某种向量排成了整齐划一的队列。很多夫妻在如何培养后代的问题上都会产生分歧，但是卡罗尔和我开心地发现我们在有关我们自己孩子的每一件事上几乎都有相同的看法。而如果我们中有一个人不那么确定，只要跟另一个人谈谈，总会恍然于心。

在这个庞大、疯狂且常常令人恐惧的世界里，抚养我们的孩子安全、幸福、智慧地长大，这个共同目标成为我们婚姻的支配性主题，并以相同的模具重铸了我们两人。虽然我们是可以截然区分的个体，但是就为人父母而言，这种区分性似乎消逝不见，甚而近乎完全消失了。首先是在生活的这一个领域之中，然后又慢慢地推及其他的领域，我们成为拥有两个身体的一个个体、一个独一的"偶人"、一个"不可分的可分人"、一个单一的"双体人"。我们两个成了偶我。我们有完全一致的感觉和反应，我们有完全一致的梦想与忧惧，完全一致的希望与恐惧。那

些希望与梦想并非分别专属于我或者卡罗尔，而后复制了两份——它们是一套希望与梦想，它们是我们的希望与梦想。

我并不想让自己听起来神秘兮兮，似乎在暗示我们共同的希望漂浮在独立于我们大脑的缥缈的梦幻岛之上。我们的希望当然要在物理上出现两次，在我们各自的大脑中分别有一次——但是如果我们从一个足够抽象的层面上看，这些希望就是一个相同的模式，只不过是在两个不同的物理介质中出现了而已。

在两个不同的细胞、两种不同的组织中，可以存在一种“相同的基因”。没有人会跟这个观点过不去。可是，基因是什么？基因不是一个实在的东西，因为如果它是的话，它就只能在一个细胞或一个组织中存在了。不是这样的。基因是一种模式——一种特定的核苷酸序列（写在纸上的编码形式通常是从由四个字母构成的字母表“ACGT”中抽取的字母序列）。所以，基因是一种抽象物，因而在不同的细胞、不同的组织乃至相隔数百亿年的生命组织中，才可能存在“完全相同的基因”。

在两种不同的语言、两种不同的文化中可以存在“同一种小说”。没有人会跟这个观点过不去。可是，小说是什么？小说不是一串特定的词语序列，因为如果它是的话，它就只能由一种语言写作且存在于一种文化之中。不是这样的。小说是一种模式——一种包括了人物、事件、气氛、情调、玩笑、典故等等的特定集合。所以，小说是一种抽象物，因而在不同的语言、不同的文化乃至相隔数千年的文化中，才可能存在“同一种小说”。

同理，也不应该有任何人跟下面这个观点过不去：在两个不同的人的大脑中，可能栖息着“同样的希望与梦想”，特别是，这两个人在一起生活了很多年，而且作为一对夫妻，还创造出了被这些希望与梦想所围绕的新的生命。这听起来可能太过浪漫了，但这就是我当时的感觉，也是我现在仍旧不变的感觉。如此多的共享，特别是关于我们孩子的共享，把我们的灵魂以某种不可诉诸行迹却又实实在在发自内心的方式联

合在一起，并在生活的某些维度上把我们变成了一个单一的结构，作为一个整体而行动。这很像一群鱼，它们也是作为一个单一思维的更高层级实体在行动。



---

1. See [Barwise and Moss].
2. See [Hofstadter 2001].
3. 这是心理学家威廉·詹姆斯对“混沌的意识流”的形容。——译者注
4. 分别是第二人称单数的第一格、第三格和第四格。

第16章
与最深之谜的搏斗

> 你会返航，重新找到生命的平衡，焕然一新，但是需要一点时间来治愈。当你回来的时候，我们都在这里，都在岸边等你。
>
> ——丹尼尔·丹尼特

一场随机的事件会改变一切

1993年12月，我的妻子卡罗尔突患脑瘤病逝。当时，我们正在意大利的特伦托（Trento）学术休假，一年的休假刚开始不到三个月。一切突如其来，完全没有任何征兆。她去世时还不满43岁，而我们的孩子，丹尼和莫妮卡才分别只有5岁和2岁。我在结婚之前，从没想象过自己有一天会如此崩溃。那双眼睛的深处闪烁着一颗光辉明亮的灵魂，而转瞬之间，那颗灵魂便黯淡了下去。光熄灭了。

至今为止，我感受到的最沉重的打击，都不是我个人的损失（“噢，我现在该怎么办？我在茫然怅惘时，谁能给我安慰？长夜难捱时，谁能给我拥抱？”）——而是卡罗尔的个人损失。我当然很想她，我对她的思念让我肝肠寸断——但是，远比这更让我不能释怀的，是她的损失：抚养她的孩子长大成人、看着他们发展出各自的人格、欣赏他们的天赋、在他们悲伤时送上安慰、为他们读睡前故事、给他们唱歌、对他们稚气的玩笑报以微笑、涂刷他们房间的墙壁、用铅笔在他们的壁橱外侧标记他们的身高、教他们骑自行车、陪他们外出旅游、向他们展示其他语言、给他们买一条宠物狗、招待他们的好朋友、带他们去滑雪

和滑冰、在我们的游戏室里一起看老电影等等。所有这些都曾被我们那么轻易地视为在未来能够水到渠成之事，卡罗尔却在须臾之间丢失殆尽。我无法接受。

过了好几个月，当我回到美国之后，我曾在一段时间里尝试过参加一个新近丧偶者的集体治疗会——“治愈心灵”，我记得他们是取了这个名字，而我发现，大多数失去了另一半的人都把焦点放在他们自己的痛苦之上，关注的是他们自己的损失，以及他们自己接下来该怎么办。当然，这个治疗会之所以起这个名字，意义正在于此——你应该治愈，应该变好。但是，卡罗尔怎么才能治愈呢？

我真的感觉，好像这个治疗会里的其他人跟我完全不在一个频道上。我们根本就没有共同的烦恼！只有我的伴侣是在孩子还很小的时候就去世了，而这个事实似乎让一切都不同了。卡罗尔被剥夺了一切，而我难以忍受想象——而又无法不去想象——她被夺走的东西。卡罗尔遭受的残酷的不公命运让我感到难以承受，而我的朋友们则不停地对我说（古怪的是，他们是抱着安慰我的好意）：“你不能替她感到难过！她已经走了！你为她感到难过的那个人已经不在了！”我感觉他们大错特错，错得离谱，错得彻头彻尾。

有一天，当我凝视着一张卡罗尔在去世几个月前拍摄的照片时，我看着她的脸庞，看得如此深入，竟至于我感到自己站在了她眼睛的深处，而忽然之间，泪水涌了出来，我听到自己说：“那是我！那是我！”简单的几个字唤起了很多我之前闪过的念头：我们的灵魂融为了一个更高层级的实体，我们两人在各自灵魂的核心寄予我们的孩子以相同的希望和梦想，这些希望并不是孤立的、各自的希望，而是一个整体的希望，一种定义了我们两人的清晰的东西，将我们熔铸成一体，那种我在结婚生子之前只能隐隐约约想象的一体。就在那时，我终于意识到，虽然卡罗尔已经不在了，她的核心部分并没有消失，而是在我的头脑中笃定地继续活在世间。

绝望的云雀

卡罗尔突然离世的悲剧发生后，我度过了充满了超现实感的几个月。她的意识忽然消失之谜萦绕在我的心头，久久不散，让我难以接受，而且我还一直觉得她仍然在我身边，这也让我困惑不解。我想把这些说不清道不明的事情写下来，可是又对自己毫无信心。于是从1994年3月下旬起，我与在麻省的密友和同事丹尼尔·丹内特开始了跨洋的电子邮件通信，因为丹关于思维和“我”的概念的观念似乎总能与我保持在几乎一致的频段上（我们在1981年合作编辑《心我论》（*The Mind’s I*）这本书时，合作十分顺畅、愉快，可能原因正在于此）。丹同样也把职业生涯的绝大多数时间都花在了思考和写作这类问题之上，所以严格地说，他算不上一个随机挑选的通信对象！

从这番通信开始，我们隔着大西洋，零零散散地来回交流达数月之久，最后一封信是我在那年8月末发出去的，发完这封信，我就和孩子们一起回国了。我们的通信相当一边倒，大概90%的“话”都是由我来说的，竭尽全力把那些难以捉摸的、有时近乎不可说的观念清楚地表达出来，而丹则主要进行简短的评述，告诉我他是否同意我的话，以及理由如何。

我在写《我是个怪圈》的最后几章时，又重新阅读了我们的整套通信，如果全部打印出来，大概有35页纸那么多。虽然它并不是什么优美的散文，我却深觉其中某些部分值得以某种形式收入这本新书。当然了，我的思索都是极其个人化的。它们是一个丈夫在自己的妻子毫无预兆地化作一股青烟而散后，于深度的震颤之中展开的殊死搏斗。我决定把这些断章收在这里，并不是因为我想就我对我妻子的爱发表某种事后的高调公开宣言——即便我无疑是爱她的，而且爱得很深。我决定收入我的这些思索，原因很简单，只是因为它们是与构成这本书核心的问题展开搏斗的发自内心的求索。我在关于人类灵魂和人类意识这个话题上所写过的一切，都不如我写给丹的那些信那样发自内心。而且，即便我

觉得自己现在对于这个问题的理解多少比当时更加清楚了，可是我仍然怀疑今天的我还能不能写出像当时那样充满急迫性的东西。那可是一段极其痛苦和烦恼的日子。

既然我在电子邮件中展开的搏斗与本书的其他部分风格迥异，而且它们写于一个不同的时期，因而我决定，我要为它们专辟出单独的一章——也就是本章。在准备阶段，我通读了35页的电子邮件，里面的内容常常杂乱无章、冗余累赘、含糊其词，还零落地散布着与话题关系不紧甚至无关的片段，于是我把它们编辑了一下，砍到了相当于原来四分之一的长度。我还重新排列了我的信件顺序，并偶尔对我筛选出来的段落进行了细小的改动，好让它在逻辑上显得更通顺。因此，你在此处看到的并非我们二人对话中来自我这一方的原始誊录稿，因为那读起来真的会相当费事；你看到的是把最重要的话题忠实浓缩后的成果。

虽然这本来是一次对话，但是我把丹的声音从这里抹去了，因为正如我在上文所说，他主要是为我炽热而情绪化的探索提供了一个平稳而沉静的甲板。我没有试图提出任何新的理论；他只是在聆听，给予我朋友的陪伴。不管怎么说，在1994年4月的一次通信中，丹曾以诗意的笔触描述了我在那段日子里经历的一切，而我觉得他的文字正好可以作为本章的一段精彩序篇，所以我还是要在下面引用他的这些文字。在这之后，所有（轻微修改过的形式）引用的内容就全部是我的声音了，均来自1994年3月到8月之间我在电子邮件中所做的冥想。

> 在缅因州，有一艘古老的竞速帆船，出没在我航行的海面附近。每当在起航线上看到它，我就很欢喜，因为它大概是我见过的最美的帆船了；它的名字也很美，叫作“绝望的云雀”。你现在正在绝望的云雀上航行，而这也是你目前应当做的事。你的感想，属于一个在我们美好的地球上遭遇并释放了生命之力的人。你会返航，重新找到生命的平衡，焕然一新，但是需要一点时间来治愈。当你回来的时候，我们都在这里，都在岸边等你。

* * *

“卡罗尔”这个名字于我而言，所指的远不仅仅是一个身体。那个身体如今已经不在了。它的所指之物，更像是一个庞大的模式，是一种风格，是包括记忆、希望、梦想、信念、爱情、乐感、幽默感、自我怀疑、慷慨胸襟、同情心等在内的一系列东西。那些东西在某种程度上是可以分享的、客观的、可以多重实例化的，有点像一个磁盘上的软件。而我对于记忆的执意书写、留下她身影的众多录像带以及我们所有人脑中关于卡罗尔的全部记忆，使她那些模式的东西依然存在，尽管存在于一个铺散开的形式之中——铺散在不同的录像带、不同的朋友和亲人的大脑、不同的陈年日记本之间。不管怎么说，在这个物理世界中，仍有一种铺散开的卡罗尔的模式清晰可辨。而在这个意义上，卡罗尔还活着。

说“卡罗尔还活着”，我的意思是，即便是从未见过她的人，也能想象接近她、围着她以及同她在一起是什么感觉——他们能体会到她的机灵、看见她的微笑、听到她的声音和笑声、听闻她年轻时代的历险、了解她与我相遇的故事、观看她与自己的小宝宝嬉戏玩耍……

可是，我仍在不断努力，试图弄清楚，我到底在多大程度上相信，因为我和其他人还有关于她的记忆（不管是在我脑海中还是写在纸面上），所以卡罗尔的意识，她的内在，还有一部分存留在这个星球上。我是一个意识非中心化的坚定信仰者，深信意识可以散开分布，所以我更愿意认为，虽然每个个体的意识首要居于一个特定的大脑之中，它总还会通过某种方式在其他的大脑中出现。因此，当处于中心的大脑遭到毁坏以后，生命个体的碎片仍在——也就是说，仍然活着。

我还相信，外部的记忆是构成我们个人记忆的一个十分现实的部分，所以我认为，哪怕只是在我随笔记录卡罗尔机智妙语的那些小纸片上，也隐约残存着一丝卡罗尔的意识，而还有一块多少更大一点儿的她的碎片（虽然还是很小），藏在我的黄线笔记本里。在

过去几个月的悲伤日子里，我在这个本子上不知写下了多少我们二人的共同经历。当然，那些经历早已在我自己的大脑中编码过了，可它们的外化将让它们有一天被其他认识她的人分享，从而以某种方式在很小的程度上，让她“复活”。因而，哪怕是写在纸面上的静态的陈述，也能包含一个“活生生”的卡罗尔和卡罗尔的意识的元素。

* * *

所有这一切让我不禁想起，在我的父亲去世几周后，发生在我和我的母亲之间的一次对话。她说，她时不时地会去看一张父亲的照片，她喜欢那张照片，照片里的父亲在微笑，而她会发现自己也在对“他”，或者是“它”报以微笑。她对于自己这种反应的评价是：“对着那张照片微笑真是太傻了，因为那<u>不是他</u>——它只是一张扁平的、毫无意义的纸片。”然后，她就会变得非常沮丧低落，甚至感到她失去他这件事更让自己悲痛难当了。我听到她痛苦的讲述之后，沉思了一会儿，虽然我能明白她的意思，但在我看来，真实情况远比她所说的更加复杂。

没错，从表面上看，这张照片就是一张死气沉沉的、没有活力的、没有灵魂的纸片，但是它到底还是<u>以某种方式</u>抓住了她，触动了她。这让我联想起那套印有肖邦钢琴曲全集的无生命也无灵魂的纸片。虽然它们只是纸张而已，但是对全世界的人都产生了不可思议的影响。所以，我爸爸的那张照片也可能有同样的作用。当我看到这张照片时，它显然可以在我的大脑中引发深层的回响，而在我的妹妹劳拉和很多其他人脑中也会发生同样的事。对于我们来说，那张照片不仅是一个具有质量、尺寸和颜色等等属性的物体，它还是一种灌注了神奇的触发力量的<u>模式</u>。

当然，除了某人的照片或者作品全集之外，还有很多其他包含了灵魂碎片的精细模式——例如，想象我们用几个小时的录像带记录下巴赫演奏风琴并谈论他的音乐的过程，詹姆斯·克拉克·麦克

斯韦（James Clerk Maxwell）谈论物理并讲述他发现光其实是电磁波的那个时刻，普希金朗诵他自己的诗作，伽利略讲述他如何发现了木星的卫星，简·奥斯汀解释她如何构思她的人物角色及他们各自复杂的私情……

我们可以拥有一种模式，可能是一整套录像带，也可能是一本长篇的日记（像安妮·弗兰克的日记一样），由此相当于拥有了一个人的大部分，包括他们的自我、他们的灵魂、他们的“我”、他们的意识和他们的内在；只不过，这个“临界物质”到底在哪里呢？如果你愿意承认，一个人的大部分存在于这个波谱（spectrum）上的某一点处，并可为我们提供了足够大的模式，那么在我看来，你也就必须承认，即使是一种比这小得多的模式，比如一张照片或者我珍爱的卡罗尔“机智妙语”收藏，已经可以提供这个真实的人（不管多么微小）并非为零的片段了——这是“由内而外的视角”，而不仅仅是在他们身边的感觉。

* * *

莫妮卡3岁的生日到了——出于显而易见的原因，这一定是一个既快乐又悲伤的场合。我带着孩子们，跟几位朋友一起到科格诺拉（Cognola）村庄的室外比萨店聚餐。我们所在的山腰就在特伦托上面不远处，视野开阔，四下高山环绕，风景宜人。小莫尼卡坐在她的高脚椅上，和我正好面对面坐在餐桌旁。因为这是一个相当令人动情的场合，卡罗尔肯定也非常想参与进来，所以我在用力地“替卡罗尔”看莫妮卡，然后我当然会开始纳闷，我到底在做什么，我产生这种想法到底是什么意思。

“替卡罗尔看莫妮卡”的想法唤醒了我脑海中一段生动的记忆。那是1983年的夏天，过去的道格和过去的卡罗尔（或者如果你不愿意这么说的话，也可以说是“年轻的道格和年轻的卡罗尔”）坐在他们在布卢明顿最喜欢的中餐馆“锅”（Wok）的露天阳台上，凝视着一个两三岁的黑发小女孩，这个可爱的小姑娘穿着一件海军

蓝的灯芯绒裙子，不停地走来走去。我们那时还没有结婚，甚至都还没有提过结婚的话题，但是我们已经经常在充满热情地讨论孩子了，而且我们两个都打心底里渴望成为这样一个小女孩儿的父母。即便没有明言，这也一定是一个我们共享的心愿。

时至今日，11年过去了，我们的女儿莫妮卡已经真的存在，而我终于能够代替过去的道格体验他曾在1983年梦寐以求的欢乐了吗？我现在可以“替过去的道格”看他的女儿莫妮卡吗？（或者我的意思是不是说“替他看我的女儿”？或者二者都是？）而如果我能够正当合理地称自己有能力替道格这么做，那么我替过去的卡罗尔这么做又有何不可？说到底，我们在多年前的夏夜渴望共同抚养一个女儿，那是一个深度共享的渴望，完全是同一个渴望，在我们两个人的大脑中同步地点燃。所以我要问的问题就是，现在的我能否代替过去的卡罗尔体验那份快乐，现在的我能否代替过去的卡罗尔看一看莫妮卡呢？

于此至关重要的是对于灵魂阐释的深度——认识到共享的目标会促生共享的特质。例如，卡罗尔一直抱有一种很深很深的愿望，希望莫妮卡和丹尼在成长过程中能成为彼此最好的朋友，而且在长大成人后也保持不变。这个愿望也存在于我的内心，并且一直都很强烈（事实上，我们一直以来都拥有这个共同的希望，而且我甚至在她去世前就已经开始竭尽全力促成它的实现了），而它现在比以往对我的行为的影响更强了。正因为她已经不在了，而我又是她在这个世界上的最佳代理人，所以我比任何时候都要更加对她负责。

* * *

除了卡罗尔的愿望和理想，她自己关于“我”的个人认识也重现在我的脑中，因为我跟她太近了，因为我跟她的共情太深了，我跟她一起感受了太多的事情，所以当我们说话的时候，我可以从她的视角出发来看待事物，不管是她生理上的痛苦（肠镜检查一个小时之后，她的体内残留的气泡不停翻滚，痛得她死去活来）、她最

开心的欢乐（莫大伟的一句魔性的机智妙语[注]，在剑桥的一顿美味的印度餐）、她最热切的希望、她看电影时的反应还是随便任何感觉。

在我们聊天中的某些简短时刻，甚或在感情强烈的凝噎瞬间，我就是卡罗尔，正如有时候，她就是道格。所以她的“个人胚芽”——借用斯坦尼斯拉夫·莱姆（Stanislaw Lem）小说《不服神役》（Non Servium）[注]中的术语——在我的大脑中形成了一个多少有点模糊的自我副本，在我的大脑中创造了第二个哥德尔式的旋涡（第一个当然是我自己的自我旋涡）。这个哥德尔式的旋涡使我得以成为她，或者换一种说法，这个哥德尔式的旋涡使她的自我或她的个人胚芽得以在我的硬件上（以一种简化的方式）运行。

但是，如今在我的大脑中存在的这第二个旋涡，这个模拟的个人胚芽，与那个曾在她的大脑内存在而如今已不在的本体旋涡，那个原始的旋涡，可以相提并论吗？卡罗尔的意识是否还存在于这个世界的某个角落？换句话说，我有可能“替卡罗尔”来看莫妮卡，甚至在最微小的程度上成为卡罗尔来看莫妮卡吗？还是说，那个个人胚芽终归还是彻底被抹去了？

＊＊＊

一个人就是一个视角——不仅是一种物理的视角（在宇宙中一个特定的物理位置上的特定眼球中所见），更重要的是一种心灵的视角：根植于一座庞大的记忆银行中的一套一触即发的关联。后者可以被另一个人吸收，随着时间越吸越多。这跟循序渐进地习得一门外语是一个道理。

在初学外语的一段时间里，一个人口中所说的话大体上都是“赝品”——也就是说，他们是在用自己的母语思考，只不过很快替换了词汇而已，所以给人留下的印象仿佛是在用那第二门语言思考一样；可随着一个人第二语言的经历不断丰富增长，新的语法

习惯形成并和数千个词汇一起，慢慢转变成反射，于是第二语言就变得越来越根深蒂固，也越来越货真价实了。人们渐渐成为流利使用另一种语言的思考者和说话者，而即便口音不正，也不再是“赝品”了。通过另一个人的灵魂来观察这个世界，也要经历同样的过程。

例如，我的父母在他们近50年的婚姻生活中，已经把彼此的心灵视角深深地内化了，因而每个人都逐渐成为另一半的“流利的身为者”。可能当我的母亲“身为”我的父亲时，她带有自己的“口音”，而反之亦然。但是，对于他们两人中的任何一人而言，身为另一人的行为都是货真价实的，绝不是假冒伪劣。

跟我的父母一样，当卡罗尔还活着的时候，我也有某种程度的身为卡罗尔，反之亦然。虽然我花了几年的时间才学会“身为”卡罗尔，而且显然从未达到“母语者”的水平，但是我觉得，公平地说，在我们最亲密的那些时光里，我算得上是我妻子的“流利的身为者”。我和她共享了那么多的记忆，不但有我们在一起的时光，还包括我们遇见之前的日子，我认识那么多塑造了她的人，我们有那么多同样热爱的音乐、电影、书籍、朋友和笑话，我还分享了那么多她最为私密的理想和希望。所以，她的视角、她的内在、她的自我，最初只存在于一个大脑中，现在又到了第二个大脑中，虽然后者远不如本来的那个完整和精密。（事实上，早在她遇到我之前，她的视角就已经产生了其他的实例，因为它显然会以不同的程度和不同的忠实水平被她的兄弟姐妹和父母内化。）不消说，卡罗尔的视角至今为止，总是在她的大脑中得到了最强的实例化。

这些有关“身为”某个别人的讨论让我想起，在20世纪70年代末的一场语言学系举办的圣诞晚会上，卡罗尔和我的老朋友汤姆·厄恩斯特（Tom Ernst）奉献了一出对后者导师约翰·金史密斯（John Goldsmith）教授的模仿，惟妙惟肖地还原了约翰（也是我的好朋友）举手投足间的种种细节。我看着汤姆“穿上”和“脱下”约翰的画风，有种毛骨悚然的感觉——汤姆在这么做的时候，把约翰穿在了身上，又干净利落地把他脱下来。

* * *

人有较浅的方面，也有较深的方面，是那些深的方面为浅的方面灌注了真正的意义。我猜这听起来可能有点隐晦。我的意思是如果我相信陈述X（例如，“肖邦是一位伟大的作曲家”），而另外某个人也相信X，那么虽然我们在表面上达成了一致，可是当我们在想到X时，我们的内在感觉可能完全不同，尽管我们在浅显的口头层面上拥有“相同”的信念。反过来说，如果我们的灵魂有一种深刻的相似性，那么我们关于X的信念就很相像了，而我们也会直觉地与彼此产生共鸣。（至少在这个话题上）交流将成为一件不费吹灰之力的事。

要想让两个人达成相互理解，真正要紧的是下面这些内容：对音乐有相似的反应（不仅共享喜欢的音乐，也共享不喜欢的音乐）、对人有相似的反应（再说一遍，还是要把喜欢和不喜欢的都包括在内）、有同等程度的同情、诚实、耐心、多情、勇猛、野心、竞争意识等等。这些构成人格、性格和性情的核心组件对于相互理解起到的决定性的作用。

例如，想一想那种不断感觉自己低人一等的崩溃体验。有些人对此感触很深，而有些人则完全没有意识。拥有强大自信的人就是*永远*也无法感受缺乏自信的那种无力感——他们“就是不明白”。而正是*这些*方面，这些灵魂的最内在的方面（与其相对的是那些相对客观且可以传递的东西，比如参观过的国家、读过的小说、拿手好菜、熟知的历史等等），才让灵魂变得独一无二。

一个人更深的方面带来了一个自我，一个“我”，而我关心的是，这些方面是否也可以传递给另一个人，或者被另一个人吸收（也就是被第二个人的大脑吸收）。那第二个人不一定要为吸收第一个人而必须改变他们自己的人格或意见；它可以更像是一个第二自我，跟一件衣服、一副人格面具或者一个舞台角色一样，可以偶尔穿戴或偷偷溜进去（我脑中浮现的画面是汤姆•厄恩斯特穿上和脱下约翰•金史密斯的人格的情形，但这里涉及的层面自然远为更

深），是一种看待世界的“第二有利视角”。

可是关键的问题在于，不论你如何吸收另一个人，你能吸收到*那么多*吗？多到让你在那个原本的大脑死亡后，还能感觉到那个*人*并没有完全从地球上消失，因为他们（至少是他们的一个典型片段）仍在你的大脑中展现，因为他们依然活在“第二个神经家园”。

* * *

在我看来，要想正面处理这个问题，人们真的必须关注那个被我称为“哥德尔式自我旋涡”的东西。于是那个关键的问题就变成了这样：当“自我”——通过一生的锁定与自我稳定过程而产生了一个“我”的那些结构——的*指针*以某种不完整的、低分辨率的方式复制到一个第二大脑中时，它们最终指向了何处？

在我内部的卡罗尔模型与（位于她自己大脑中的那个）原初的自我模型相比，肯定更加“瘦削”或稀薄，但那种稀薄性并非此中枢要。真正的关键之处在于：即使在我内部的卡罗尔模型无比丰富（比如，像我的母亲所有的我的父亲的模型，或者甚至比那还强上10倍），就引发一个“我”的目的而言，它会不会仍旧是一种*错误的结构类型*？它是否并不是一个怪圈？这个结构是不是并*没有*指向*它自身*，而是指向*别的东西*，因此缺乏那种制造一个“我”所必需的回转、涡旋、自我指涉的品质？

我的猜想是，如果这个模型极其丰富、如实，那么其中所有指针的指向实际上都会是*可变的*——换言之，在我的卡罗尔模型的内部的指针可以滑动，既可以指向她在*我的*大脑中的符号，也可以同样有效地指向她*本人*的自我符号。真若如此，那个原初的旋涡，那个原初结构的“我”，就已经成功地传递到了第二种介质之内，并在其中如实地重建起来（虽然模糊了很多）。

构成自我“外”层的大量指针，主要指向世界的标准而普遍的

方面（比如，雨水、冰激凌、燕子掠过等等）；构成自我“中间”层的指针更多地指向与自己的生活联系更紧密的东西（比如，一个人父母的面孔和声音、热爱的音乐、从小长大的街道、童年时心爱的宠物、最喜欢的书和电影以及很多其他深刻的事物）；最后，内部的圣所则有无数混成一团的指针指向很深的“索引式”（indexical）之物，如一个人的不安全感、性欲、最强烈的恐惧、最深沉的爱，还有其他很多我说不真切的东西。所有这一切都说得十分含糊，目的只是为了提出一幅图景，其中最外层有最多指向外部的箭头，中间层混合了指向外部和指向内部的箭头，而最深处的内核则有无数箭头直向自我指回。这是一座“怪圈之城”——也是为你准备的一个“我”！

据我猜想，最深处的内核具有的那种深度扭转回自身的品质，令其太难传递到别处了，正因如此，灵魂才会与一个单一的身体和一个单一的大脑如此紧密地依附在一起，而且几乎不可逆转。当然，相对缺乏向内指的指针的外层，传递起来也相对简单，而中间层的传递难度则是中等。像我跟卡罗尔关系这么近的人，可以得到大量的外层、某些部分的中间层和聊胜于无的内核，但是到底能否有人内化足够多的内核，以至于哪怕在最弱的意义上敢于说出：“她依然活在我们中间”？

＊＊＊

可能我是把传递的难度放大了。在某种意义上，所有的哥德尔式自我环路（即引出了一个“我”的怪圈），在大致水平上都是同位素，因此在最低的近似标准下，它们的传递也许一点儿都不难；把它们与彼此区分开来的只是它们的“风格”，这包括了记忆，当然还包括遗传倾向和天赋等等。所以，我们可以成为变色龙，输入其他人生命历史的“成分”（这些成分带着独一无二的个体性注入了他们的自我环路），仅就这种程度而言，我们的确有能力通过他们的眼睛观看世界。他们的心灵视角是可以传递和模块化的——并

没有仅仅困在一个容易消亡的硬件内部。

如果这是真的，那么卡罗尔就还活着，因为她的视角留存了下来——或者不如说，她的视角留存了下来，就此意义而言，她还活着，不仅在我的大脑中，也在其他人的大脑中。这就是记录的好处，这就是为什么我们要把记忆写下来，拍照、录影，而且留下的记录越清晰越好——因为在这些记录的帮助下，你可以“占有”或“被收入”其他人的大脑。这也正是为什么弗雷德里希·肖邦，那个真正的人，直至今日仍然如此生动地活在我们的世界中。

* * *

总有一天，我会观看那些记录了卡罗尔影像的录像带，到那个时候，我一定会心碎不已，因为我将再次看到她，再次体验到她的生命，再次和她厮守在一起——而虽然我将被爱意包围，可我还是难以抑制地觉得这些都是假货，我在被戏弄，而所有这一切又会让我纳闷，在我的大脑里到底发生了什么事。

观看那些录像带，会在我的大脑中激起种种模式——自她去世以来，我大脑中的这些信号还是第一次被触发、再次激活、复苏并重现生气，它们将在我的内部翩翩起舞。毫无疑问，这些如今仅仅是记录在画面中的模式，同她本人在场并真正在做录像带中的那些事时在我的大脑中激发的模式相比，在强度上没有差异。在我大脑内部由视频触发的符号舞蹈，也将是同样的舞蹈，而且跳舞的也是同样的符号，丝毫无异于她就站在我的眼前。

所以，在我的大脑里就存在了这样一套结构，视频、照片以及其他情感极其强烈的记录可以通过一种特别深刻的方式通达这些结构——这些在我内部的结构，当卡罗尔活着的时候，与她关联在一起，与她形成深度的共鸣，并表征了卡罗尔，而且无论如何，这些结构看起来就是卡罗尔。但是当我观看录像时，鉴于我知道她已经不在了，立刻就会有一种虚假感迎面而来，同时也让我陷入深深的困惑，因为我将会仿佛真的看到她，仿佛真的复活了她，仿佛真的

把她唤了回来，就像我在梦里做的那样。所以我便会思考，在我的大脑中共同形成“卡罗尔符号”的那些结构，本质上*到底*为何物？卡罗尔符号究竟有多大？而重中之重是：道格内部的卡罗尔符号，距离*成为*一个人，而不仅仅是*表征*或者*用符号表示*一个人，到底有多近？

下面这个问题应该简单很多（虽然我不认为它真的更简单了）。当J.D.塞林格创作《麦田里的守望者》时，他大脑中的“霍尔顿·考尔菲德”符号的本质是什么？那个结构就是霍尔顿·考尔菲德拥有过的全部了——可是它却非常非常地丰富。那个符号可能不如一个完整的人类灵魂那么丰富，但是霍尔顿·考尔菲德看起来*太*像一个人了，有一个真实的内核、一个真实的灵魂、一个真实的个人胚芽，尽管那只是一个“微观的模型”。不管在塞林格的大脑中构成霍尔顿·考尔菲德符号之物究竟是什么，你都不能要求一个人在另一个人大脑中有比它更加丰富的表征或镜照了。

* * *

丹，我希望你能听出来，这里的一整套观念有一以贯之之处，虽然我之前说出来的显然都是些支离破碎的思路。要想说明白这些事情实在是太难了，特别是在深切情绪的干扰下，更是难上加难。情绪*希望*事情按照想要的特定方式存在，在某种程度上推动答案偏向希望的一侧。当然，这些问题之所以能达到如此紧张而强烈的地步变得如此重要，恰恰来自那些欲望的力量。如果悲剧未曾降临，这一切都不会发生。

我必须承认，我感觉自己有点像一个在量子力学发展的期间与量子力学的现实展开搏斗的某个人[注]，只不过是在量子力学还没有完全严格地建立之前——这个人大概活动在1918年左右，是个像索末菲（Sommerfeld）一样的人，他对于那时可用的全部所谓“半古典”模型（美妙的波尔原子及其改进版本）都有深刻的理解，但是此时距海森堡（Heisenberg）和薛定谔（Schrödinger）登上历史

舞台、切入问题的核心并排除一切困惑还有一段时间。在1918年左右，许多真理都近在咫尺，可即便是一只脚踩在边沿的人，也很容易退回纯古典的思考模式，并陷入无望的困惑之中。

这正是今天的我对自我、灵魂和意识的感觉。我感觉自己仿佛已经相当接近认识到意识的可分散性和灵魂的错觉了，可还是不能击中要害。我明明知道自己思想深处的观点在根本上是违反直觉的（“量子力学”），可又能感到自己在这些问题上不断地滑回惯性的直觉（“经典”）观点，这实在是太令人沮丧了。

* * *

又及

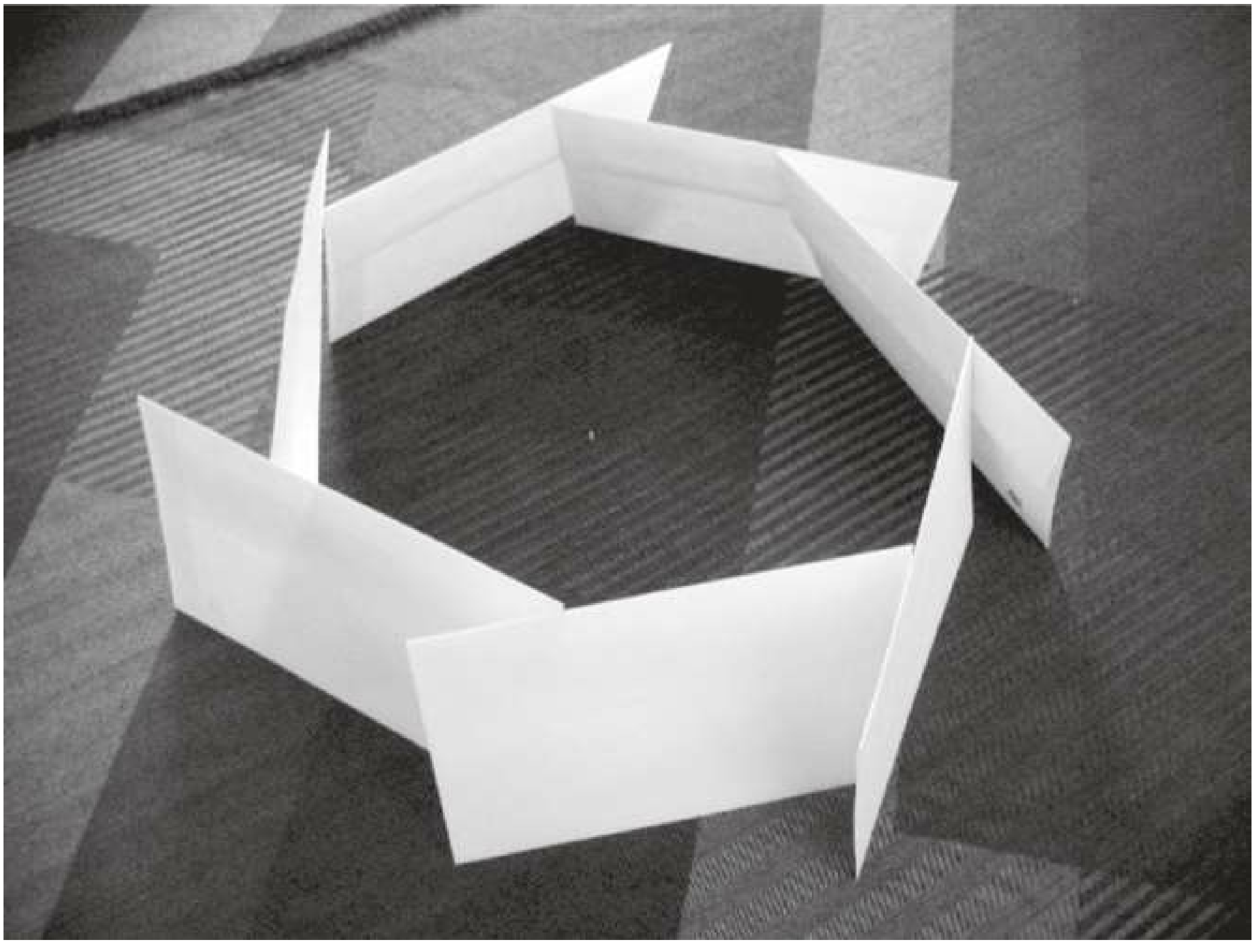
在本章的内容整理完成之后很久，我忽然想到，可能有些读者会忍不住想要总结出，随着卡罗尔之死[注]，她极其忧伤的丈夫在丧偶的可怕重压之下发生了某种扭曲，开始动手构建某种精雕细琢的形而上学，想借此自欺欺人否认在所有局外人的眼中不证自明之事：他的妻子已经死了，彻底消失了，就是这样。

这种怀疑论乃至犬儒论都是相当正常的，而我也要承认，即使是我，当回头再看这些搏斗时也禁不住怀疑这些关于灵魂和存活的痛苦思索，到底有没有很大一部分动机是来自我对于死亡的现实性或终结性的抗拒；我不只是说1994年的思索，还包括那之后很多年。我对自己是相当了解的，因而我不会真的认为情况如此（虽然我有时候也会有点儿不太确定情况到底是怎么样的）；可是，一想到并不了解我的读者可能轻易地得出这样的结论，进而把我的搏斗贬低成一个深受苦难之人情绪使然的胡言乱语，是他为了抚慰自己的悲伤而对自己信念系统的权宜篡改，我就难免会感到相当的困扰。

因此，当我最近在我文件柜里重翻到一些旧文件时——那些文件的名字都是“同一性”“怪圈”“意识”等等，我不免松了一口气。那些文件中的大量笔迹白纸黑字地记录了所有之前谈到的观点，而那些文字远在地平线上的阴影尚未露面之前就已经存在了。我发现自己纯手写的天马行空的思考，讨论的是人类灵魂模糊的特质，特别是，我还发现有几段文字明确地讨论了有关卡罗尔和我的灵魂融合成紧密的单一结构之事，还有卡罗尔和丹尼的“灵魂融合体”。

在那些即兴写就的篇章里，我常常构想出相当有趣却又十分严肃的思想实验，我在实验中调整两个大脑之间可能的信息流动的速率（其中有一次涉及在我的大脑和一个僵尸大脑之间建立的联系——至少在我看来，这是个相当讨人喜欢的想法）。显而易见的是，这些关于我们是谁以及一个人何以独一无二的观念已经在我的脑海中酝酿了几十年，而当我结婚之后，尤其是当我体验到生子以及同某个人一起抚养他们长大，而那个人对他们的爱又与我对他们的爱如此相似、如此难分难解地纠缠在一起时，所有这一切都激烈地沸腾起来了。

我的书已经写完了，而那些旧文件则成为它丰满的前身。也许某一天，我在那时候写过的某些东西会得见天日，也可能它们永远都不会发表，但至少我很欣慰自己可以得知，我在最艰难的日子里并没有顺势倒向对我暗送秋波的某种阻力最小的信念系统，而是始终坚守自己在多年前用心找出的原则，未忘初心。关于自己这一点的认识，给我带来了小小的慰藉。



1. 在我跟卡罗尔结婚后不久的一天晚上，我们邀请几位朋友到我们在安娜堡的家里吃印度菜。梅兰妮·米歇尔和莫大伟明知卡罗尔做印度菜的糟糕厨艺，还是欣然赴约。结果就在最后时刻，我们已经年逾八旬的最老的客人忽然打电话给我们说，他们不太能应付很辣的食物，这把卡罗尔的下厨计划一下子都打乱了。尽管如此，卡罗尔临危不乱，华丽转身，做出了一顿完全不同却相当美味的大餐。晚饭用毕，主客相谈甚欢，两三个小时之后，客人纷纷告别，最后只留下了大伟、梅兰妮和我们夫妇二人。我们又聊了一会儿，然后就在他们准备告辞之前，卡罗尔不经意地提到了今天本来要准备的菜肴，还解释了为什么她没有按照原来的食谱准备。转瞬之间，莫大伟佯装出巨大的愤慨，暴跳着大喊：“为什么，你们这些说好了做印度菜又不兑现的人（Indian-dinner givers），你们！”

2. See “Non Serviam” in [Hofstadter and Dennett], which is a virtuosic philosophical fantasy masquerading as a book review (of a book that, needless to say, is merely a figment of Lem’s imagination).

3. [Pais 1986], [Pais 1992], and [Pullman] portray the transition period between the Bohr atom and quantum mechanics, while [Jauch] and [Greenstein and Zajonc] chart remaining mysteries.

4. See Chapter 15 of [Hofstadter 1997], another place where I discuss many of these ideas.

第17章
我们如何活在彼此之中

通用机器

1993年，在我12岁左右的时候，市面上出售一些小工具配件，你可以买回家，把它们组装成可以执行各种有趣功能的电路。你可以组装出一台收音机、一条可以加二进制数的线路、一件可以使用一套替代密码来编码和解码信息的设备、一个可以跟你玩井字格游戏的“大脑”以及其他诸如此类的装置。这些机器中的每一种都是专用的，也就是说，它只能玩一种把戏。这也正是我们在成长过程中形成的有关“机器”的普遍意义。比如，一台电冰箱是为了冷藏而专用的机器，一个闹钟是为了叫醒而专用的机器，我们对这些观念早已习以为常。可是时至今日，我们已经渐渐开始习惯遇见超越其原始目的的机器。

以手机为例。如今，为了提高自身的竞争力，手机的卖点已经不再那么（也许甚至极少）依赖于它们作为通信设备的原始目的了，相反，手机比的是它里面备有多少铃声、可以玩几种游戏、像素有多少，天知道还有哪些其他的东西！手机曾经是专用的机器，但现在已经不是了。为什么呢？因为它们内在的线路已经超越了某个特定的复杂性阈限，令它们得以具有如变色龙一样的本性。你可以使用手机内部的硬件装载文字处理器和网页浏览器，还有一堆乱七八糟的游戏，等等。实际上，这正是计算机革命的全部要义：当某个明确定义的阈限——我把它称作“哥德尔-图灵阈限”——被跨越之后，一台计算机就可以模拟任何一种机器了。

这便是“通用机器”这个词的含义[注]。1936年，英国数学家和计算机先驱阿兰·图灵（Allan Turing）最早提出了这个术语；到了今天，我们对这个术语的基本观念已经再熟悉不过了，只不过大多数人并不知道它在专业上的术语或概念。我们已经习惯了从互联网上下载各种软件，把我们的笔记本电脑转变成临时的专业化设备，用来看电影、听音乐、玩游戏、打便宜的国际长途电话，反正没有做不到的，只有想不到的。各种各样的机器通过电线为我们所得，甚至通过无线以软件或模式的形式涌入并栖居在我们的计算机硬件里。一台通用机器，一拍脑门儿就能变化出新的功能，或者更准确地说，只需要双击一下鼠标。我在我的电子邮件程序、文字处理器、网页浏览器、照片显示器和10多种其他“应用”之间来回跳转，而它们都住在我的电脑内部。在任意一个特定的时刻，这些相互独立的、专能专用的机器大多都在休眠、沉睡、耐心地（事实上是无意识地）等待被我尊贵的鼠标双击唤醒，然后服帖地跳出来，听从我的指挥。

哥德尔把*PM*映射到了它自己的身上，而阿兰·图灵正是在他的启发下，意识到这种计算机通用性的关键阈限之精确所在，即一台机器能否灵活到足以阅读并正确解释一套描述其自身结构的数据。在这个关键的枢纽处，一台机器原则上可以一目了然地看到它如何一步一步地完成任何一项特定的任务。图灵意识到，一台机器如果灵活性达到这种关键水平，就可以模仿其他任何一种机器，不论后者有多么复杂。换句话说，没有任何东西比一台通用机器更加灵活。走到通用性，就是走到头了！

这也就是为什么，只要我给我的苹果电脑安装上合适的软件，它就能跟我儿子那台更贵、更快的“外星人”电脑（不管是运行什么特定程序）表现得毫无差别，反之亦然。唯一剩下的差别就是速度了，因为我的苹果电脑在内核深度上永远只是一台苹果电脑。因此，它要想模拟高速的外星人硬件，就必须不断访问明确描述外星人硬件的数据表，而做所有这些查阅是相当缓慢的。这就好像是我想让你学我的签名，于是努力写出一长串指令，告诉你怎么画出每一段微小的曲线。原则上这是可

能的，但这比我用我自己的手直接签名要慢上百倍。

通用性的意外

在这种通用的机器和我在前文描述《数学原理》的力量时说到过的通用性（虽然我没有使用这个词）之间，有一种紧密的类比性关联。伯特兰·罗素和阿尔弗雷德·诺思·怀特海未曾料及，而库尔特·哥德尔则意识到，仅仅借助对正整数某些根本特征的表征（比如交换律、分配律、数学归纳法），他们已经在无意中让他们的形式系统*PM*跨越了一个关键的阈限，令其具有了“通用性”（universality），也就是有能力定义那些模仿任意复杂程度的其他模式的数论函数（或者切实地说，甚至有能力转过来模仿自我——引出了哥德尔的黑带招式）。

罗素和怀特海没有意识到他们所创造出来的究竟为何物，因为他们根本没有想过要用*PM*去“模拟”任何其他的事物。那个想法干脆就不在他们的雷达屏幕上（说起来，那个时候，你在任何人的雷达屏幕上都看不到雷达本身）。质数、平方数、两个平方数之和、两个质数之和、斐波那契数等，都仅仅被视为美妙的数学模式——这些模式虽然精妙繁复且无限迷人，仍是由数字构成的模式而已，人们不会想到它们可能与任何其他东西同构，更别提作为其他东西的替代物去表示那些东西了。可是，等到哥德尔和图灵出现之后，这种天真幼稚的想法便瞬间告灭，付诸东流了。

大体上，设计出最早的电子计算机的工程师们跟罗素和怀特海一样，对于他们无意中创造的丰富性也是毫无察觉。他们以为他们制造的机器只适用于十分有限的、纯粹军事上的范畴——例如，把风向和空气阻力考虑在内，计算飞行导弹的发射轨迹的机器，或者破解特定的敌军暗号密码的机器。他们把他们的计算机视为目标单一的专门化机器——有点儿像上了发条的音乐盒，每只盒子只能播放一首曲子。

但是等到某一天，当阿兰·图灵主要基于哥德尔1931年文章而提出的抽象计算理论与具体的工程机械现实碰撞在一起时，更具观察力的人们（特别是图灵本人和约翰·冯·诺依曼）把眼前的事实联系起来，意识到他们的机器包含了整数运算的丰富性，而哥德尔已经展示过这种丰富性的力量有多么强大，因而他们的机器也便拥有了通用性。忽然之间，这些机器好像变成了可以读取任意一条打孔纸带的音乐盒，可以播放*任何一种*曲调了。从那时起，到手机可以开始扮演很多种角色，而不再只限于简单的移动电话的功能就是迟早的事情了。它们只需要跨过那个把它们局限在单一“曲调”上的复杂性和存储容量的阈限，然后便可以成为任何东西了。

早期的计算机工程师只把他们的计算机当作数字运算设备，而并没有把数字看作一种通用的媒介。今天的我们（我所谓的“我们”是指我们的文化整体，而不只是专家）仍然不会那样看待数字，但是我们在理解上的欠缺却是出于一个完全不同的原因——事实上，是出于完全相反的原因。如今，因为所有的数字都在我们的笔记本和电脑桌面背后隐藏得太好了，所以我们才完全忘记它们的存在。我们在我们的电脑屏幕上观看虚拟的橄榄球游戏，而对战双方是只存在于（根据设计原理执行算术指令的）中央处理器的“梦之队”。孩子们建造起虚拟的小镇，镇上小小的虚拟居民虚拟地骑着虚拟的自行车飞驰而过，虚拟的叶子从虚拟的树上飘落，虚拟的炊烟飘散在虚拟的空气中。宇宙学家创造出虚拟的星系，放任它们自行运转，并观察它们在虚拟的相撞之后会发生什么。生物学家创造出虚拟的蛋白质，并观察它们如何根据它们虚拟的亚分子成分的复杂的虚拟化学性质发生折叠。

我可以列举出上百种发生在电脑屏幕上的这类事情，但是极少有人曾经想过这样的一个事实，即所有这些事情之所以能够发生，都有赖于在底部的硬件层面上*整数的相加与相乘*。我们把电脑叫作*计算机*，毕竟不是白叫的！事实上，它们就是在计算以二进制表示的整数的和与乘积。而在这个意义上，哥德尔在1931年打开的那个眩晕了世界、摧垮了

罗素并颠覆了希尔伯特的洞见，在我们如今这个充斥着下载、升级与千亿字节的文化中已经太过习以为常了，乃至于虽然我们无时无刻不浸入其中，却几乎没有人对此有一丁点儿的意识。那个原初的洞见，在我们的身边仅剩下一丝可见的踪迹，或者不如说是“可听见”的踪迹，而那正是“计算机”一词。如果你愿意拨冗三思，这个词可以提示你：事实上，在所有那些五颜六色的图像、令人欲罢不能的游戏以及快速的网页搜索的底部，真正发生的只有整数的运算。这是多么滑稽的笑话呀！

实际上，事情远非如此一目了然，而所有的原因，我都已经在第11章详细说明过了。但凡存在一种模式，它要么可以被看作其自身，要么可以被视为代表了任何与其同构之物。用在小美离家出走这件事上的话，同样也可以放在奥蕾丽的身上，即使其中某一种情况是说这些话的本意，两种解读方法依然具有同等的真实性。同理，一次二进制整数的运算（比如，把“0000000011001111”转变成“1100111100000000”），第一个人可能把它描述成与256相乘，第二个人则可能把它描述成小数点向左移8位数，第三个人则可能把它描述成一种颜色从一个像素移动到另一个像素之中，而第四个人则可能把它描述成删除一份文件中的某个字母或数字。只要每个人都正确描述了发生之事，就没有哪个人有特别的优势。我们把电脑叫作“计算机”的理由已经是过去时了。它们起源于整数运算的机器，如今继续如此描述它们当然依旧是有效的——但是正如库尔特·哥德尔在1931年首次意识到的那样，我们如今也已经意识到，用一种与它们的发明者最初意图完全不同的美妙方式观察和讨论这些设备也是同样有效的。

通用生命

我们人类也是一种通用机器：我们的神经硬件可以复制任意的模式，尽管生物进化从未定下“表征通用性”的伟大计划。先后通过我们的感觉以及我们的符号，我们可以内化各种不同种类的外部现象。例如，

当我们观看涟漪在池塘水面上荡开时，我们的符号会映现它们的圆形形状，把它们抽象化，并可以在很久以后重放那些形状的精华。我之所以要说“精华”，是因为有些——其实是大部分——细节丢失了；如你所知，我们不会记住我们所见之事的所有层面，而只会保留那些我们的硬件在自然选择的压力下认定为最重要的层面。我还不得不澄清（虽然我希望没有读者会掉入这个陷阱），当我说我们的符号“内化”或者“复制”外部模式时，我的意思并不是在说，当我们观看一片池塘里的波纹时，或者当我们“重放”关于这一场景（或者很多类似场景模糊在一起）的一段记忆时，有真正的圆形模式在我们大脑内部某个水平平面上散开。我的意思是说，有一组与水、湿润、池塘、水平平面、圆环形、扩散、上下冲撞的东西等等相联系的结构被联合激活了。我说的可不是头脑里的一个电影屏幕！

表征通用性还意味着，我们可以理解观念与事件，而无须身临其境。例如，我在第11章提到过，人类（与绝大多数动物不同）能够轻松地处理电视机屏幕上的二维像素矩阵，而且可以把那些不断变化的矩阵视为随着时间演变的某种遥远或虚构的三维场景的编码。

当我和我的孩子们前往离家很远的内华达山脉去度滑雪假期时，我们把家里的金毛猎犬奥利送到布卢明顿狗舍寄养。多亏了互联网的技术，我们通过“狗狗摄像头”抖动的静止摄影看到了20多只狗在围栏围住的户外场地漫无目的地闲逛的画面，真是让人大饱眼福。那些狗有点儿像正在进行布朗运动的粒子，虽然每一只在画面上都只由相当小的一块像素序列来呈现，我们还是经常能够通过尾巴角度等微妙的细节认出我们家的奥利。出于某种原因，我和孩子们都觉得这种偷窥奥利的行为特别有趣。然而，虽然我们可以把这个古怪的场景讲述给我们的人类朋友听，虽然我敢赌一大笔钱说，仅仅是上面这几行字，就足以在你的脑海中唤起一幅关于狗舍的狗和在滑雪胜地的人的生动画面，可是我们都意识到要想给奥利解释我们在数千英里之外“监控”它的事，是根本没有希望的。奥利永远都不会知道，也永远都不可能知道。

为什么不能呢？因为奥利是一条狗，而狗的大脑并不是通用的。它们不能理解诸如“抖动的静止摄影”、“24小时摄像头”、“监控狗舍里的狗”乃至“2 000英里以外”这样的观念。这是横亘在人类与狗之间的一个巨大而根本性的鸿沟——实际上，这条鸿沟也存在于人类与任何其他物种之间。正是这条鸿沟把我们人类区分出来，让我们独一无二，最终赋予了我们那种被称为“灵魂”的东西。

在生物的世界里，当一个系统的符号集合可以不受任何显在限制地延伸扩展时，表征通用性的魔法阈限便被跨过了。这条阈限挡在早期灵长类动物与我们自己之间的某个物种等级之处。位于这个相当于哥德尔-图灵阈限之上的系统——我们暂且简称其为“生命”——拥有在他们自身内部模拟它们遇到的其他生命的能力，可以为与自己短暂相逢的生命快速地拼凑出草率而简陋的模型，也可以慢慢地把粗糙的模型加以改进和细化，甚至可以一步到位地创造出想象中的生命。（对于偏好创造其他生命的生命，通常有一种不太正式的叫法，那就是“小说家”。）

一旦越过了那个魔法阈限，通用的生命似乎不可避免地生发出一种对于品尝其他通用生命内在的贪婪渴求。这就是为什么我们会发明电影、电视剧、电视新闻、博客、网络摄像头、八卦专栏、《人物》杂志以及《世界新闻周刊》。人们渴望进入其他人的头脑中，从别人的大脑里“向外看”，拿他人的经历大快朵颐。

虽然我的描绘带着一点儿冷嘲热讽的调调，但是表征通用性及其创造的那种对于他者经历的近乎难以满足的渴求，距离同情只有一步之遥；而同情，正是在我看来第一可敬的人类品质。以一种深刻的方式“成为”另一个人，并不仅仅是在知识上用他们的方式来看待世界，也不仅仅是扎根于塑造他们成长轨迹的时空中去体验和感觉；远远不止于此。“成为”另一个人，是要接受他们的价值观，承担他们的欲念，抱持他们的希望，感受他们的渴求，分享他们的梦想，畏怕他们的恐惧，参与他们的生活，与他们的灵魂融为一体。

有人来访

不久前的一个清晨，我醒来时，脑海里充满了关于我父亲的记忆。在一闪而过的瞬间，我的思绪似乎在半梦半醒间把他唤回了现世，尽管“他”必定只是悬浮在我大脑舞台的稀薄介质之中，却是那么活灵活现。不论如何，在短短的一段时间里，那种感觉就好像他是真的回来了，然后忽然间，悲伤袭来，他噗的一下消失了。每个成年人对于这种苦甜参半的体验都再熟悉不过了，可是我们应该如何理解这种体验呢？这些居于我们内部的软件生命到底拥有多大程度的真实性呢？为什么我在几行之间，要把“他”放在引号里呢？为什么要如此小心翼翼？为什么要竖起藩篱？

当你不止一闪念地梦到或者想到了某个你心爱的人（不管那个人是已经去世多年还是正在跟你打电话）时，究竟真正发生的是什么？使用本书中的术语体系，则此中发生之事便毫无歧义可言了。那个人的符号在你的头骨内部被激活了，从休眠状态中惊醒过来，就像你双击了电脑桌面上的一个图标那样明确无误。而这发生的瞬间，就像一个电脑游戏在你的屏幕上打开一样，你的思维也开始表现出与在“常态”语境下不同的行为。你允许你自己被一个“外来的通用生命”入侵，而且在某种程度上，这个外来者掌控了你的头骨内部，开始以它自己的方式指东指西，制造话语、观念、记忆，而且平常不会存在的联系也在你的大脑内部冒了出来。心爱之人符号的激活转而引发了一整套代表那个人珍爱的风格的相应倾向，以及它们嵌入世界之内和向外观察世界的独特方式。结果就是，在你大脑之内，你会给自己带来不断的惊喜，想出你在正常情况下编不出来的笑话、带着一种不同的情绪看待事物、做出不同的价值判断，不一而足。

可是对于我们而言，眼下的关键在于这个问题：你所拥有的另一个人的符号真的是一个“我”吗？那个符号能有内在体验吗？或者说，它是不是跟你所拥有的一根木棍、一块石头或者操场上一架秋千的符号一

样，也是没有生命的呢？我选择操场上的一架秋千来举例是有原因的。就一架秋千而言，不管你把它安放在哪个操场上，不管在你的想象中它的座位是用什么材料做的，也不管你把吊着它的横梁想象得有多高，你总能看到它前后晃来晃去，看到它以秋千的有趣的扭动方式轻微地扭来扭去，看到它如果没有推力就会渐渐失去能量，还能听到它的链子在轻柔地叮当作响。虽然没有人会说秋千是有生命的，但毫无疑问的是，它的思维替代物却能在你大脑滚烫的底座上翩翩起舞。说到底，这正是一个大脑存在的意义——为活跃的符号提供一个舞蹈的平台。

我在本书中一直反复论述，概念是大脑中活跃的符号(注)，如果你像我一样，打心底里相信这一点，同时也真心相信，跟其他认知对象一样，人在大脑中也是通过符号得到表征的（换句话说，一个人认识的每一个人，都由大脑中的一个概念对其进行了内化的镜照，尽管这会是一个非常复杂的概念），而且最后还笃定地相信，一个自我也是一个概念，只不过是一个更加复杂的概念（即一个“我”、一个“个人胚芽”、一颗坚实的“玻璃球”），那么这套信念必然引发的一个不可避免的结论就是，你的大脑也在不同的程度上容纳了其他的“我”、其他的灵魂，其程度取决于你对它的表征有多么忠实，以及你与目标个体形成了多么深刻的共鸣。我把“形成共鸣”这个附加条件也加入进来，是因为没有人能随心所欲地滑入任何一个成熟的灵魂，正如一个人不能穿进每一件旧衣服一样；某一些灵魂和某一些服装，就是比另外的选择更“契合”。

化学反应及其缺乏

对我而言，阐述两个灵魂之间的契合或“共鸣”有好有坏这个观念的最佳范例，就是音乐品位了。我永远都不会忘记30多年前发生的那件事。一位钢琴家朋友对贝拉·巴托克（Béla Bartók）的《第二小提琴协奏曲》推崇备至，并坚持要我去听一听。她这是一次投桃报李，因为我曾

在几年前把据我所知最激情澎湃的乐章之一推荐给了她——普罗科菲耶夫的《第三钢琴协奏曲》。那一次，她对于普罗科菲耶夫的最后一段乐章产生了令人难以置信的剧烈共鸣，这个事实似乎表现出我们两个的音乐品位比较契合；因此，我把她对于巴托克《第二小提琴协奏曲》的热情背书相当重视。出于怂恿我的目的，她说巴托克不仅多次反复使用了她最爱的普罗科菲耶夫和弦，而且用得比后者还要更好。不用多说了！我马上走出去买了一张碟片回家。那天晚上，抱着很高的期望，我把碟片播放出来，用心倾听。但令我颇感失望的是，我完全没有受到任何触动。这真是太让人不解了。我又听了一遍。然后又一遍。又一遍。又一遍。在两周的时间里，我一直在听那首受到高度赞扬的曲子，即便没有20多遍，也不下10多遍，可是它在我的内心仍然没有激起任何波澜，除了中间15秒左右的一段稍稍地打动了我。你可以说这是我内在的一个盲点——或者一个聋点——或者按照我更愿意用的说法，你可以干脆地说，我的灵魂与巴托克的灵魂之间的“契合”是极其贫乏的。而这一点已经在多次与巴托克音乐的遭遇中反复得到了印证，所以现在我已经相当确信，当我听巴托克的时候，我的内在会发生什么（或者不如说是没有发生什么）。虽然我也喜欢他创作的一些（基于民歌改编的）短乐章，但是他大部分重头的作品都是我根本欣赏不来的。因此，我对于这位朋友和我之间存在很多音乐上的共通之处这种感觉也被大大削减了，事实上，连我们的友谊也自此逐渐地淡漠了。

在写完上面一段之后，我越来越好奇，30年前的一段记忆有没有可能不太可靠，或者我的灵魂是不是已经在这段时间里打开了全新的音乐视野，所以我径直走向我的唱片播放器（没错，是黑胶唱片机），再一次播放了巴托克的《第二小提琴协奏曲》，从头到尾用心听了一遍。我的反应和之前是一模一样的。对我而言，那首曲子就好像是在逛来逛去，漫无目的，不得要领。听着这段音乐，我感觉像是一个磁场砸在了一个超导体之上——连一微米都穿不透！未免这个隐喻太过晦涩，让我们不如干脆说，我突然止步不前了，就停在了最表面的一层上。我根本摸不到它的头绪；它是一种无法理解的音乐风格。听这音乐，就好像在

看一本用外星文写的书。你能看出来它的背后是有智力的——也许有很多！可是你完全不知道它讲的是什么。

我重提这件有点消沉的往事，因为它代表了生命中的千万种经历，因为没有更好的词可用，我们不妨把这些经历中所包含的东西叫作人与人之间的“化学反应”。巴托克和我之间就是没有化学反应。我敬重他的智力、他的创造力以及他高尚的道德标准，可是我完全不晓得让他心动的东西是什么。一点儿线索都没有。但是我可以说，成千上万的人都有如此的感觉——可也有很多的人感觉与此恰恰相反。比方说，在这个世界上，没有任何一曲音乐比普罗科菲耶夫《第一小提琴协奏曲》之于我的意义更为重大，而它的创作与巴托克的《第二小提琴协奏曲》相隔只有寥寥数年。事实上，让我大惑不解的是，我甚至看到过有人把这两首乐曲相提并论，仿佛它们是从一个模子里面刻出来的。它们也许在某些地方具有表面质地上的某些共通性，但是在我看来，它们之间的差异堪比巴赫与埃米纳姆（Eminem）。巴托克从我的左耳朵进，从我的右耳朵出，而普罗科菲耶夫则像一剂妙用无穷的万灵药水，流遍我的全身。它对我说话、在我的体内飞腾、点燃我的激情、把生命的音量调到爆裂。

我不需要再唠叨下去了，因为我相信每个读者都体验过这一类的化学反应和化学反应的缺乏——甚至可能是在巴托克和普罗科菲耶夫《第一小提琴协奏曲》上与我恰恰相反的体验；可即便如此，我想要努力传达的信息也并不会因此而褪色，反而将依然清晰而鲜明。在我看来，音乐似乎是通往心灵或者心灵之间的一条直接路径——事实上，是最直接的路径。音乐品位的全盘投合，包括喜爱与憎恶——后者是某种极少遇见的东西——是我所发现的对于灵魂亲密关系最为确切的指南。而灵魂之间的亲密关系则意味着，关系中的人能迅速地了解对方的本性，而且具备活在彼此之中的巨大潜力。

通过吸收熔化的流星而生长的“拷贝”星

不管是儿童、青少年还是成人，我们都是盲目的拷贝者（copycat）。我们不自觉地、自动地把各种各样其他人的行为碎片收入我们自己的仓库。我已经提到过我上小学一年级时模仿过的“卡西迪牛仔般的微笑”，我想它至今仍若隐若现地活跃在我“现实”的微笑之中。我还有另外几十种有关模仿他人动作的明确记忆，都发生在那个年纪以及之后的日子里。我还记得自己羡慕并模仿过一个朋友东倒西歪、凹凸不平的字体，一个趾高气扬的同学夸夸其谈的耍酷姿态，一个学长大摇大摆的走路姿势，电影《环游世界八十天》中法国售票员说“**américain**”（法语“美国”）这个词的发音方式，一个大学朋友在每次通话结束时都要说一次对方的名字的习惯，等等。而且当我观看有我在内的录像视频时，我总是猝不及防地看到那么多专属于我妹妹劳拉的熟悉表情（它们太像她了）在我的脸上一闪而过。这些表情到底是谁从谁那儿借来的？是什么时候？为什么？我永远都不可能知道答案。

我长久以来一直在看着自己的两个孩子模仿他们美国朋友朗朗上口的发音模式和最爱用的短语，还能听到某些特定的意大利朋友的声音和短语回响在他们的意大利语当中。有时候，在听他们其中的任意一个人说话时，我竟能跟着那些单词和发音，不假思索地列出他们朋友的名字。

我曾经抱着强烈的热情——那热情强烈到仿佛就是纯粹的我——创作过一些钢琴短曲，可颇具讽刺意味的是，那些乐谱却充满了可以识别出来的他人的特征，这些特征明显来自肖邦、巴赫、普罗科菲耶夫、拉赫玛尼诺夫（Rachmaninoff）、肖斯塔科维奇（Shostakovich）、斯克里亚宾（Scriabin）、拉威尔（Ravel）、福雷（Fauré）、德彪西（Debussy）、普朗克（Poulenc）、门德尔松（Mendelssohn）、格什温（Gershwin）、波特（Porter）、罗杰斯（Rodgers）、科恩（Kern）以及很轻易就能举出来的很多其他作曲家，我曾在那些年里浸润在他们的

音乐之中。我的写作风格也带有数不清的作家的印记，他们驾驭文字的美妙方式是我梦寐以求想要模仿的。我的观念来自我的母亲、我的父亲、我年轻时代的朋友们和我的老师们……我做的每一件事都是某种对从他人那里借来之物的改写，那些人要么真的与我很亲近，要么就是与我有一种虚拟的亲近感，而虚拟的影响最为深远。

我的质地的大部分是由从成百上千个人的经验中借来的零星碎片编织而成的，那些名人与我素未谋面，而且几乎肯定不会跟我遇见，所以他们对于我而言只是“虚拟的人”。下面是抽取出来的一份样本：尼尔斯·波尔（Niels Bohr）、苏斯博士（Dr. Seuss）、卡洛尔·金（Carole King）、马丁·路德·金、比莉·哈乐黛（Billie Holiday）、米奇·曼托（Mickey Mantle）、玛丽·马汀（Mary Martin）、马克辛·沙利文（Maxine Sullivan）、安瓦尔·萨达特（Anwar Sadat）、查尔斯.德内（Charles Trenet）、罗伯特·肯尼迪（Robert Kennedy）、保罗·狄拉克（P.A.M.Dirac）、比尔·科斯比（Bill Cosby）、彼得·塞勒斯（Peter Sellers）、亨利·卡蒂埃-布列松（Henri Cartier-Bresson）、朝永振一郎（Sin-Itiro Tomonaga）、杰西·欧文斯（Jesse Owens）、格劳乔·马克斯（Groucho Marx）、珍妮特·玛戈林（Janet Margolin）、罗尔德·达尔、弗朗索瓦兹·萨冈（Françoise Sagan）、西德尼·贝彻（Sidney Bechet）、雪莉·麦克雷恩（Shirley MacLaine）、雅克·塔蒂（Jacques Tati）和查尔斯·舒尔茨（Charles Shultz）。

上面提到的所有人都对我的人生产生了重大的积极影响，而他们的人生与我所处的时代都有一部分的重叠，因此我（至少在理论上）是有可能与他们本人相遇的。但是我的内部还包含了数不清的来自千百个我绝不可能遇见或与其交往之人的踪迹，比如W. C.菲尔兹（W. C. Fields）、伽利略、哈里·胡迪尼（Harry Houdini）、保罗·克利（Paul Klee）、克莱芒·马罗（Clément Marot）、约翰·巴斯克维尔（John Baskerville）、胖子沃勒（Fats Waller）、安妮·弗兰克、霍尔顿·考尔菲德、尼摩船长（Captain Nemo）、莫奈、莱昂哈德·欧拉、但丁、普希

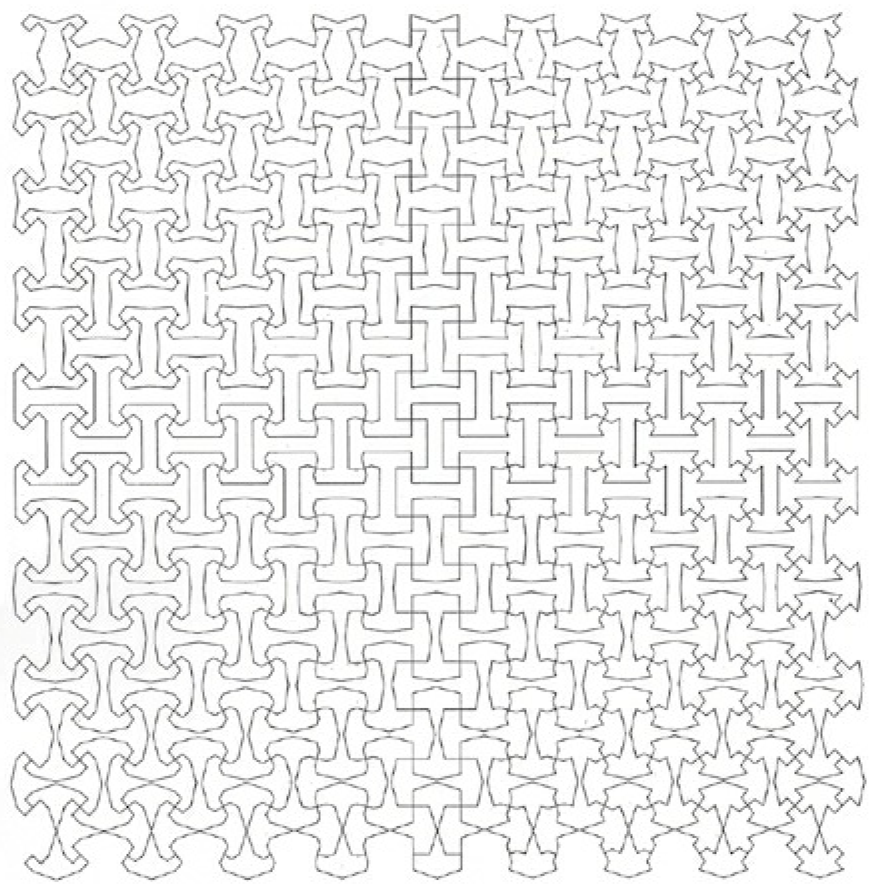
金、尤金·奥涅金、麦克斯韦、匹克威克（Samuel Pickwick）、查尔斯·巴贝奇（Charles Babage）、阿基米德和查理·布朗。

当然，后一份清单中的人有一部分是虚构的，而其他人介于现实与虚构之间，但是这都不重要，在我的思维中，他们都仅仅是虚拟的生命。不管是虚构/非虚构还是虚拟/非虚拟的维度，都不要紧，要紧的是一个个体与我的内在交流的时长与深度。就此而言，则霍尔顿·考尔菲德与普希金大概处于同等位置，都远远高于但丁。

我们都是奇妙的拼贴画，是古怪的小小的“拷贝”星，通过累加其他人的习惯、观念、风格、怪癖、玩笑、习语、语调、希望和恐惧而成长起来，仿佛它们都是从天而降、忽闪而至的流星，与我们相撞并结合在了一起。最初，这种不自然的、外来的举止缓慢地融入我们的自我之中，好像蜡在日光下熔化，然后逐渐成为我们的一部分，跟它们曾经作为他人的一部分的状态没有什么不同（虽然那个人也很可能是从别人那里借来的）。虽然我的流星隐喻听起来仿佛我们是随机相撞的受害者，但我并不是想暗示说，我们会欣然接受任意一种闯入我们星体表面的行为习性——我们是挑三拣四的，通常只接受我们羡慕或渴求的那些特征，但是即使是我们的选择风格，其本身也会随着时间的推移受到影响；作为反复累加的结果，我们会渐渐转化成另外的样子。曾经在星球表面上正当存在的东西，会像古罗马的废墟一样，逐渐被掩埋起来，随着我们的半径不断扩大，越来越接近我们的地心。

以上这一切都在指出，我们每个人都是一束他人灵魂的碎片，只不过是以一种新的方式组合起来而已。但是，灵魂碎片每一位贡献者的表征当然不是平等的。那些我们爱的人和爱我们的人，是在我们内部获得最强大的表征之人，而我们的“我”也正是通过他们所有人的影响在很多年里复杂的相互冲撞而形成的。戴维·奥尔森（David Oleson）在1964年创作过一幅绝妙的“拼花地板变形”钢笔画[注]（见上页），不仅用画面阐明了这个观念，而且还暗藏了一个巧妙的双关，因为它的题目是“我在

中心”：



我们在这张图的中心可以看到一个隐喻的个体，其形状（字母“I”）是由所有邻近形状构成的结果。而它们的形状也同样是由他们的邻近形状构成的，如此类推。越接近这幅设计图案的边缘，图形之间彼此的差别就变得越来越大。这是多么绝妙的一个视觉隐喻啊！我们人类同样是由与我们接近的人所决定的，特别是那些离我们最近之人！

一个人能吸收多少另一个人的内在？

当我们在一家商店的收银台与一位收银员打交道的那几分钟里，我们显然没有构建出关于那个人内在热情的精细表征。我们对于她的表征

偏颇不全、浮光掠影，以至于几天之后再见到这个人时，我们甚至都认不出来了。当圣诞节购物狂潮来临，我们走在拥挤的人行道上时，与我们擦肩而过的成百上千人，每个人都会落入与那位收银员同样的情况，有过之而无不及。虽然我们很清楚，每个人在他们的核心处都有一个跟我们多少有点儿相似的怪圈，但是赋予它独特性的那些细节实在是我们难以触及的，所以他们的核心方面完全不在我们的表征范围之中。相反，我们收录进来的只有浅显的方面，跟他们内心的热情、跟他们究竟是谁，都毫无关系。这种情况是我们在自己的大脑中为大多数与我们萍水相逢之人建立的一种典型的“截断式走廊”图像；我们对与他们内心中那个怪圈一无所知。

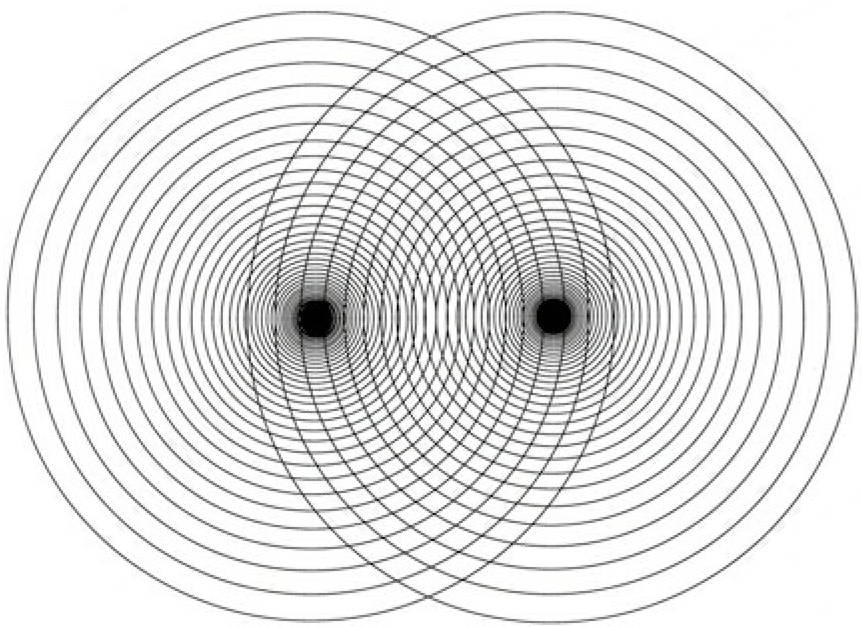
我上面列举过的很多著名人物都是占据我的特质的中心之人，意思是说，我无法想象，如果我没有听闻过他们的观念或事迹，我将会成为一个什么样的人。但是还有成千上万其他的名人，他们与我的生命仅仅是擦肩而过，有时候令人不快，有时候令人愉快。这些更加边缘的个体在我内部的表征主要是各种各样知名的成就（不管他们对于我的影响是好是坏）——说过的一句名言、发明的一个公式、抓拍的一张照片、设计的一种字体、接住的一个直线球、鼓动的一众暴民、拯救的一群难民、策划的一个阴谋、挥就的一首诗、提出的一次和平建议、草创的一本漫画、构思的一则妙语或者吟唱的一曲民谣。

相比之下，那些位于中心的人物在我的大脑中是由复杂的符号进行表征的，这些符号远远超出了他们留在身后的外在踪迹；他们在我的内部注入一道额外的微光，这道光照亮了它活在他们的大脑中的方式，以及它透过他们的眼睛向外看这个世界的方式。我感觉我已经进入了他们内在的隐藏领地，在有些情况下进入得很深，而他们反过来也渗入了我的内在。

可话又说回来，不管我们最爱的那些作曲家、作家、艺术家等人给我们带来了多少美好的影响，我们难免还是与自己真正熟识的人、与自

己交往多年的人以及自己心爱之人的关系更加亲密。我们对这些人的关爱如此之深，让他们达成某个特定的个人目标也成为我们重要的目标，而且我们还会花相当长的时间来思考如何实现那个目标（我特意选择了“那个目标”这个没有特指的短语，因为这到底是他们的目标，还是我们的目标，是说不清楚的）。

我们活在这些人的内部，而他们也活在我们的内部。回到那个涉及两个交互的视频反馈系统的隐喻，则与我们关系紧密的某人在我们的屏幕上是由在我们自己那条无穷无尽的走廊之外另加的第二条无穷无尽的走廊来进行表征的。我们可以一直看下去——它们的怪圈、它们的个人胚芽，都包含在我们的内部。然而，让我们再次回顾那个隐喻，既然我们的摄像头和我们的屏幕是由颗粒组成的，我们对于所爱之人的表征，无论是相比于我们自己的自我表征还是相较于他们自己的自我表征，都是不够精确的。



双击所爱之人灵魂的图标

当我在1994年通过与丹尼尔·丹尼特的电子邮件展开沉闷深思时，我曾一度很担心自己在卡罗尔去世后第一次观看记录她身影的录像时会有什么样的感觉。我想象的是，我脑中的卡罗尔符号会被录像带上的图像强烈地唤醒——比她离开以后任何时候都要强烈。而我对于这次激活将创造的那种幻觉的力量深感畏惧。我将仿佛看到她站在楼梯边上，然而，如果我站起身，穿过房间，走到她本来站过的那个地方，我就会发现那里根本就没有人。虽然我将看到她明亮的笑脸，听到她开朗的笑声，我却不能走上前去，用我的胳膊搂着她的肩膀。观看那些录像带会加剧我失去她的痛苦，因为我看起来仿佛物理性地把她唤了回来，可是在现实中，却根本做不到这样的事。录像带带不回她的物理本性。

但是，她的内在本性呢？当卡罗尔还活着的时候，她的在场总是能在我的大脑中激发特定的符号。再明显不过的是，那些录像视频将再次触发那些相同的符号，虽然不如之前那样丰满。那么，在我的大脑中被激活的符号舞步的本性到底是什么？当那些视频不可避免地双击到我的“卡罗尔”图标时，在我的内部将发生什么？从休眠的昏暗处中喷涌而出那种奇怪的、复杂的东西将会是一种现实之物——或者不管怎么样，至少跟在我内部的“我”一样真实。然后，关键问题是，在我的大脑中那个奇怪的东西与曾经在卡罗尔的大脑内部盛开的“我”究竟有多么不同？它是一种完全不同类别的东西，还是一种仅仅是精细度降低了的同类东西呢？

用另一个人的大脑思考

在1994年那个焦灼的春天，丹尼尔·丹尼特对于我的搏斗做出了很多回应，其中有一句话在我的脑海中挥之不去，印象鲜明：“从你所说的来看，很明显，卡罗尔将在未来很长一段时间里用你的大脑思考了。”我很感激他这种煽情的说法，也与其形成了强烈的共鸣。后来我发现，丹其实是在我们共同的朋友、人工智能先驱马尔文·明斯基

（Marvin Minsky）的首肯下引用了他的话——拷贝无处不在！

“她将用你的大脑思考。”丹尼特-明斯基的这句话之于我的意义大致如下。在某些特定的情境下，输入给我的信号在我的大脑中所走的路径并没有通往我的记忆，而是通向了卡罗尔的记忆（或者不如说是我对于它们低分辨率的“副本”）。我们孩子的面庞、她的父母和兄弟姐妹的话音、我们房子里的房间——这些东西在加工时而会发生在赋予了它们卡罗尔风格意义的一个参照系之中，它们所在的参照系将使它们扎根于她的经历，并与她的经历联系起来（依然是我大脑中粗糙的翻版）。在我接收的信号之上额外叠加的语义，发源自她的生活。如此一来，当达到了某种程度之后，我们多年的共同生活令我得以精确地吸收并移植那些让卡罗尔在这个地球上留下烙印的经历，那么她就能够通过我与世界互动，通过我继续生活下来。等达到了那个程度，也只有当达到那个程度时，卡罗尔才能用我的大脑思考，用我的心感觉，在我的灵魂中生活。

不同颗粒度的马赛克

既然所有的事情都悬在了“到达某种程度”这几个字之上，那么这里至关重要的似乎就是对于本体的忠实程度了。我很快就为这个观念发明了一个隐喻，喻体是由彩色小石块拼成的马赛克所呈现的肖像画。一个人了解你的程度越深，你在他们头脑里的“肖像”就越具有精细的颗粒度。分辨率最高的你的肖像当然是你自己的自画像——你本人用一生的时间为自己拼出的马赛克，即你的自我符号，颗粒度是极其精细的。由此而及卡罗尔的例子，她本人的自我符号是迄今为止颗粒度最精细的肖像，描绘了她内在的实质、她的内心之光、她的个人胚芽。而我的卡罗尔马赛克，她的内在于我大脑之内更粗颗粒的复制品，当然在分辨率上仅次于它。

不用说，我的卡罗尔肖像比她自己的肖像颗粒更粗；怎么可能不是这样呢？我没有在她的家庭中长大，没有跟她读同一所学校，也没有经历过她的童年和青少年。可尽管如此，经由我们共同度过的许多年月，通过数千小时惬意而亲密的交谈，我已经吸收了太多处于她的特质的中心地位的那些经历的低分辨率的副本。卡罗尔关于她年轻时代的记忆——她的父母、兄弟姐妹，陪伴她童年的柯利犬巴尼，全家前往盖茨堡、博物馆和华盛顿的“教育郊游”，他们在密歇根州中部一座湖边木屋里度过的那个暑假，她在青少年时期对花花绿绿的袜子的偏爱，她在青春前期对于阅读和古典音乐的热爱，以及她发现自己与那么多与她同龄孩子迥异而隔膜的感觉——所有这些都在我的大脑中留下了它们的副本，模糊不清，但总归是副本。她的一些记忆太过鲜活，简直成了我自己的记忆，仿佛我真的经历过那些日子。某些怀疑的人也许会不屑一顾地说：“只是假的记忆而已。”而我的回应是：“真和假的差别在哪儿？”

我有一位朋友曾经给我讲过他的一次观光旅行。他描述的细节栩栩如生，如映眼前，乃至过了几年之后，我以为我自己也经历了那次旅行。甚至更过分的是，我甚至都不记得我的那位朋友跟“我”的那次旅行有任何关系了！有一天，我们在聊天的时候提到了这次旅行，我们两个自然都坚持认为，我们各自才是那场旅行的主人公。那简直是真假美猴王，佛也说不清了！尽管如此，当我的朋友给我亮出了他在那场旅行中的留影并回忆起远比我记忆中更多的细节时，我意识到还是我错了——但是谁知道在我的思维里还发生过多少次没有被纠正过来的其他混淆呢？它们留下的虚假记忆，已经成为我的自我意象中不可割离的内在要素。

说到底，真实的个人记忆与虚假的记忆之间到底有什么差异呢？微乎其微。我至今仍记得小说《麦田里的守望者》或电影《大卫与丽莎》（*David and Lisa*）中的特定情景，仿佛它们都曾在我的身上发生过一样——就算它们没有，又如何呢？它们跟真正发生过一样清晰。同样的说法也适用于其他种类艺术作品中的很多情景。它们是我情绪图书馆的组

成部分，以休眠的状态储存在那里，等待在适当的触发下重焕生机，跟我“真正”的记忆并无不同。我从自己的亲身经历中回忆起来的事和我从别人的故事中回忆起来的事，没有绝对的根本性区别。随着时光流逝，一个人的记忆（以及假记忆）的锐度也遭到磨损，那时这种区别就更加模糊难辨了。

模式的移植

即使大多数读者都同意我所说的大部分内容，对于他们中的很多人来说，还是有最难理解的一点，就是我怎么能够相信，我的脑中一个符号的激活，竟能够捕获另一个人对于这个世界的一手经验，即另一个人的意识；不管那个符号有多么精密，这也是难以想象的。我到底是哪根筋坏了，才会以为别人的自我——我的父亲、我的妻子的自我——也能经历感觉？符号的激活依赖于我的大脑中的神经硬件，而另一个人大脑中的每一个细胞都早已灰飞烟灭了。

此中的关键问题，既十分简单也非常质朴：真正的硬件到底要紧吗？卡罗尔的细胞如今已经回归到我们星球上庞大的非人格化共生系统之中；可是，只有她的细胞才有可能撑起我所谓的“卡罗尔感觉”（仿佛感觉被刻上了一个让它们独一无二的烙印）吗？还是说，其他的细胞，甚至是在我内部的细胞，也可以胜任这项工作？

在我的心中，这个问题的答案一清二楚。在一个大脑内部的细胞不是意识的承载者；意识的承载者是模式。真正至关重要的是组织的模式，而不是物质材料。无关肉体，重在动作！否则，我们就不得不给我们大脑内部的分子赋予它们在大脑外部并不具备的属性。例如，当我看到一个马上要被丢掉的篮子里躺着最后一片墨西哥玉米片时，我可能会想，“噢，你这个幸运的玉米片！如果我把你吃下去，而你那些毫无生气的分子又足够幸运，能顺着我的血管进入我的大脑并在那里驻扎下

来，那么它们就可以享用作为我的经验了！所以我必须把你嚼烂吞掉，不让你那些死气沉沉的分子有机会享受成为人的体验！”我这个想法在我的每一位读者听来都是荒诞可笑的。但是，如果组成你的分子不是你的感觉的“享受者”，那什么才是呢？剩下的就只有模式了。而模式可以从一种介质复制到另一种介质之中，即使是在完全不同的两种介质之间也没问题。这样一种行为被称为“移植”，或者，用一种更简单的说法，就是“翻译”。

一本小说可以经得住移植，尽管“目标语”的读者并没有在来源语的土壤中生活过；关键一点在于，他们在他们自己的土壤中经历过实质上相同的现象。事实上，所有的小说，不管是翻译小说还是本土小说，全都依赖着这种可移植性，因为哪怕大家都说同一门语言，也没有任何两个人在完全一致的土壤中长大。要不然的话，我们这些当代的美国人怎么能够欣赏简·奥斯汀的小说呢？

卡罗尔的灵魂也能够移植到我的大脑所在的土壤，因为我虽然没有在她的家庭中长大，没有跟他们一起搬过好几次家，但是我在某种程度上了解她早年生活的全部关键细节。她的灵魂正是从她早年的内在之根中生长出来的，而那内在之根如今仍茁壮地活在我的内部。我的大脑的肥沃土壤与她的灵魂土壤虽然并非完全一致，却十分相似。正因如此，我才得以“身为”卡罗尔，虽然稍稍带着点儿道格的口音。这就好比普希金叙事诗《叶甫盖尼·奥涅金》的詹姆斯·法兰（James Falen）译本，那可爱、轻盈而婉转的英文翻译当然毫无疑问就是那首叙事诗，虽然带着某种美国人的口音。

当然，令人悲伤的事实是，完美的副本是不存在的，而我对于卡罗尔记忆的拷贝也不够完整和完美，完全无法接近原本的精细度。当然，令人悲伤的事实是，在我的大脑中栖身的卡罗尔已经缩小成从前的那个她的一个微小的部分。令人悲伤的事实是，我大脑中卡罗尔本质的马赛克比在她的脑中占据优势地位的马赛克粗糙了太多。这实在是令人悲伤

的事实。死亡的刺痛是难以否认的。只不过，死亡的刺痛并不像它表面上看起来那么绝对和完整。

当日食发生时，太阳四周仍有日华，围成一圈光晕。当某人逝去时，他们也在那些与他们亲近之人的灵魂里留下了一抹鲜艳的余晖。随着时光流逝，这道余光难免渐渐褪色，乃至熄灭，但那将是多年以后的事了。最终，当所有亲近之人都告别了人世，余烬方凉，直到此时，才真正迎来“尘归尘，土归土”。

几年前，我的电子邮件笔友詹姆斯·普拉斯（James Plath）得知了我关于上述内容的热切沉思，于是从卡森·麦卡勒斯（Carlson McCullers）的小说《心是孤独的猎手》（*The Heart is a Lonely Hunter*）中摘取了一段，转赠给我。我正好可以用这一段话结束本章。

> 次日，日上三竿之时，他坐在楼上的房间里缝补起来。为什么？为什么在真爱中，留在世间的人常常不会自我了断，与心爱之人共赴黄泉？只是因为活下来的人必须埋藏死者吗？因为死后的繁重礼节必须有人来做？因为独活在世的那个人仿佛一时间走上了舞台，每一秒都膨胀得无限绵长，而他还被无数双眼睛盯着？因为他有必须执行的使命？或者也可能是，在爱情中，丧偶之人必须留下，以备心爱之人重生——所以离去的那个人其实并未真正死亡，而是在活者的灵魂中成长，第二次被创造而生？

---

1. See [Hennie] and [Boolos and Jeffrey].
2. See Chapter 11 of [Hofstadter 1979].
3. For a dozen-plus examples of this subtle Escher-inspired art form, see Chapter 10 of [Hofstadter 1985].

第18章
人类同一性的模糊之光

我做他人的东道主，也到他人那里做客

“一个身体，一个人”是人类最为普遍共享的信念之一，或者换一个同等的说法就是，“一个大脑，一个灵魂”。我把这个观念称为“笼中鸟隐喻”，其中的笼子当然就是大脑，而鸟则是灵魂。这种意象实在是不言自明，而且潜移默化到我们每个人思考自我的方式之中，以至于当把它明确说出口时，听起来就好像在说“一个圆，一个圆心”或“一根手指，一片指甲”一样没有意义；而质疑它则充满风险，容易给人留下神经错乱、思维古怪的印象。话虽如此，质疑这个观念却正是前面几章的目的所在。

与笼中鸟隐喻相反，我在此提出的观念是，既然一个正常的成年人类大脑是一个表征性的通用“机器”，而且人类又是社会动物，那么在一个成年人类的大脑中，就不仅仅有一个构成了与大脑形成首要联系人的同一性的怪圈，还有很多怪圈模式，即居于其他大脑中的首要怪圈的粗颗粒副本。因而大脑1就包括了怪圈1、2、3等等，每个怪圈都有各自不同的细节水平。但是这个道理不仅适用于大脑1，而且适用于任何一个大脑，所以我们又可以反过来得出这样的结论：每个正常成年人类的灵魂都以程度不一的忠实度居住在很多大脑之中，因此，每个人类的意识或“我”都不同程度地同时生活在一套包括不同大脑的集合之中。

当然，每一个特定的“我”都有一个“首要的居所”或“主要大脑”，这就意味着，像“我的灵魂居于我的大脑之中”这样简单的常识化陈述，仍

然保留了不少真相；可是，这类陈述虽然接近真相，却缺少了某个至关重要的部分，也就是那个乍听起来可能有点奇怪的观念："我的灵魂在较轻的程度上生活在不属于我的大脑之中。"

到这里，我们至少该简短地思考一下，"我的大脑"和"不属于我的大脑"这些听起来清白无害的短语有没有特别的含义。如果我有5个姐姐，那么如果我说"我的姐姐"，就算不是完全没有意义，至少也是充满了歧义。同样，如果我拥有3种国籍，那么"我的国籍"这种说法也是有歧义的。类比而言，如果我的自我符号存在于，比方说，15个不同的大脑中（必然有相应的15种不同的忠实度），那么不但"我的大脑"这个短语具有歧义性，而且连"我的"这个词也有歧义了！说话的人是谁？我想起了湾区的一家如今已经倒闭的酒吧，我每次开车路过它时都会被它牌匾上的标语逗得前仰后合："我兄弟的地儿！"没错，但是到底是谁的兄弟的地儿呢？说话的究竟是什么人？我一直都没有搞清楚这一点（我猜，也没有任何人能弄明白），但我对这个牌匾故意的卖傻欣赏有加。

幸运的是，一个"主要大脑"的存在意味着"我的大脑"具有了一个明白无误的首要意义，哪怕说出这个短语的灵魂以较轻的程度同时存在于其他的14个大脑之中。通常情况下，说这个短语的灵魂都是在使用它的主要大脑（以及它的主要身体和主要口舌），所以绝大多数的听众（包括说话者在内）都可以毫不费力地明白它的意思。

要想找到一个能与笼中鸟隐喻对抗的强大而鲜明的隐喻绝非易事。注我饶有兴致地尝试了不少可能性，包括了各种各样风马牛不相及的事物，比如蜜蜂、龙卷风、花朵、星星和驻外大使馆。一群蜜蜂或一片星云的意象足以清晰地传达出可分散性的观念，但是其中却没有可以清晰地与笼子（或者不如说头脑、大脑或者闹壳）相匹配的对应物。（蜂巢不满足我的要求，因为一群飞舞的蜜蜂根本就没有在蜂巢内部。）单体龙卷风的意象很令人心动，因为它包含涡旋体，能让人想起我们常常谈到的视频反馈环，而且这种旋涡还有一部分在空间中四散开来，但这里

依然没有“原址”的对应物，我们也不清楚，一个单体中是否存在一个首要的龙卷风。然后，我又想到了一种植物的意象：它的茎在地下蔓延四散，并同时在不同的地点冒出枝芽；它有一个首要的主干，并生长出次生的分支，这是我这个观念的一个重要组成部分。另外，我还想到了一个国家在很多其他国家设立驻外大使馆的意象，也涵盖了我所寻求的一个重要方面。但是我对以上任何一种隐喻都不满意，所以，我没有将就着使用其中的某一个，而是把它们像这样一股脑儿地撒在这里，希望它们能在你的头脑中激发出合适的意象。

感觉到一个人在别处

所有这些关于一个人同时居于好几个身体中的讨论，看上去可能都与“常识”大相径庭。常识明明白白地告诉我们，我们总是身处一个地点，不会同时出现在两地，或者更多的地方。但是，我们不妨稍微检验一下这个常识性的公理。

如果你在一家I-Max电影院里乘坐一辆狂野的过山车，那么你在哪里？你可能想说：“我坐在电影院里”；可如果真是这样的话，你为什么要这么害怕呢？几十排静止不动的椅子、爆米花的香味和一张悬挂在四五十英尺以外的薄薄的屏幕，有什么好怕的呢？答案是显而易见的：当你在看电影的时候，你的大脑接收的音视频输入感觉上并非来自电影院内部，而是来自别处，那个地方距离电影院相当遥远，并且跟电影院一点关系都没有。而正是那种让你不由自主地想要理解的输入，告诉你自己现在身处何方。你感觉到自己被运送到了一个并非你身体所在的地方，而且如此说来，你的大脑也不在那个地方。

当然，因为看电影是一种我们非常熟悉的行为，所以我们并不会为这种虚拟错位的现象感到困扰。我们知道自己只不过是把疑心暂时悬置了起来，以此让自己得以虚拟地、间接地、反复无常地进入另一个世

界。这种经历似乎引不出什么严肃的哲学难题，可是对我来说，这道初具雏形的小小裂缝却提示着，自我同时现身多处这扇大门，还可以远为更加广阔地打开。

下面让我再回忆一下从加利福尼亚州滑雪胜地内华达山脉经由“狗狗摄像头”和互联网被传送回布卢明顿狗舍的那番经历。看着那些狗在它们的小场地游戏时，我和我的孩子们一点儿都没有感觉到我们“披着奥利的皮”，但是让我们把这个场景稍稍调整一下。例如，设想传送视觉图像的带宽极大提升。不仅如此，摄像头也不再仅仅是挂在围栏内游戏场地上方的一个固定位置，而是安装在奥利的头上，并装有一个麦克风。最后，再设想你有一副专用的护目镜（装有耳机的眼镜），只要戴上它，它就可以把场景以很高的视听还原度传送给你。只要你能随时戴上或摘下这副眼镜，这副有时空传送功能的护目镜看起来就仅仅是一个游戏而已，但是如果它们连续几个小时绑定在你的脑袋上，成为你窥见外部世界的唯一方式，又会如何？你难道不会开始逐渐有点儿觉得自己仿佛<u>就是</u>奥利吗？如果你自己的眼睛和耳朵无法为你提供任何来自加利福尼亚州这边的输入，那么你身在遥远的加利福尼亚州滑雪胜地这件事，于你而言还有什么意义吗？

你可能会反对说，如果你控制不了奥利的动作，那么你就不可能感觉自己是它。既然如此，我们可以增加一个游戏手柄，让它控制奥利向左转和向右转，完全由你决定（至于具体怎么做到这一点，没必要在这里讨论）。这样一来，你的手在控制奥利的行动，而你接收到的视听输入全部来自装在奥利头顶的摄像头，如此连续几个小时不停。这个场景的设定是有点儿古怪，但是我觉得你很容易就能发现，你很快就会开始感觉自己仿佛更像是身处印第安纳州的犬舍，可以在那里自由自在地活动，而不是在某个加利福尼亚州的滑雪胜地，基本上只能坐在自己的椅子上不动（因为你戴着护目镜，所以你看不到自己移动的场所，自然也就不敢冒险移动）。在超级快速的数据传输的支持下，你能感觉到你自己在远离你的身体和大脑的某个地方，我们把这种感受称为“远程呈

现”。这个概念在20世纪80年代左右由帕特-贡克尔（Pat Gunkel）提出，并由马尔文·明斯基普及。

远程呈现对阵“现实”呈现

我在把《集异璧》的书稿输入计算机时，感受到了有可能是我最为鲜活的一次远程呈现体验。那还是20世纪70年代的晚期，要一位作者做类似这样的事情还是闻所未闻，但是我足够幸运，拥有当时世界上仅有的两套计算机排字系统之一的使用权限。不过，那两套排字系统恰好都在斯坦福，麻烦之处在于，我当时是印第安纳大学的助理教授，工作在很远之外的布卢明顿，而且我在周二和周四都有课时安排。让事情更加棘手的是，当时并没有互联网，所以我不可能从印第安纳远程完成录入工作。为了给我的书排字，我不得不人在斯坦福，但是我的教学课表只允许我在周末的时候过去，而且还不能是每个周末。所以，每当我在周末坐飞机来到斯坦福时，都会立即冲入文图拉大厅（Ventura Hall），把自己关在故事中的“Imlac室”的终端机上，一头扎进工作，并一直维持强度极高的工作状态。我有一次曾连续工作了40个小时，直到彻底崩溃。

这一切到底与远程呈现有什么关系呢？好吧，在斯坦福的每一次冗长而艰苦的工作都具有相当高的催眠性，当我离开之后，我仍然感觉自己仿佛还在那里。有一次，当我返回布卢明顿之后，我意识到自己在某一章中犯下了一个严重的排字错误，惊惶之下，我给我的朋友斯科特·金（Scott Kim）打了个电话，他当时也一直在Imlac室内没日没夜地工作。发现他还在那里，我大松一口气。斯科特很乐意坐在一台Imlac终端前，调出正确的程序和需要操作的文件。所以我们就开工了。我用语言引导斯科特执行一整段绵长而细节繁复的过程，而斯科特则把他在屏幕上看到的文字读给我听。因为我刚刚在那里待了数不清的时辰，所以我很容易就能在我的脑海中看到斯科特转述给我的任何东西。而且我还记得，当我时不时想起自己的身体仍在布卢明顿时，我感觉到的方向迷

失感有多么强烈，因为我完全感觉自己仿佛就在斯坦福，在亲手操作Imlac终端。别忘了，这种强烈的远程呈现的视觉感受，单纯是由电话里的声音模式带来的。就好像我的眼睛虽然仍在布卢明顿，可它却正盯着加利福尼亚州的一台Imlac屏幕，这多亏了斯科特的眼睛以及他在电话里清晰的言语表达。

如果愿意的话，你大可以把我的感觉称为“错觉”，但是在你这么做之前，不妨想想这种如今看来已成古董的远程呈现的实现方式有多么原始。今天，人们可以轻松地想象把所有的科技旋钮调高数个量级。在加利福尼亚州，可以有一个行动机器人，它的动作受到我即时而精确的掌控（类似那个游戏手柄），而它的多媒介“感官”把它们收到的信号实时地传递给身在印第安纳的我。结果就是，我可以完全沉浸在来自距我的大脑几千英里之外的虚拟体验，而这种体验的时间可以无限延长。最让人感到晕头转向的仍然是那个转变的瞬间：当我摘下那只让我感觉自己人在加利福尼亚州的头盔时，我会感觉自己在飞秒间被传送回2 000英里外的东边——或者反过来，当我戴上那只头盔时，我转眼便一路飙到了西海岸。

说到底，我怎么才能推断出，我在印第安纳的在场比我在加利福尼亚州的在场“更真实”呢？我想，其中一条线索来自一个泄露天机的事实：如果我要想“人在”加利福尼亚州，总是必须戴上某种头盔才行；而如果我想“人在”布卢明顿，我根本不需要任何这种设备。另一条线索可能是，如果我在加利福尼亚州游荡时得到了某种食物，我没办法把它送到我在印第安纳的胃里！可是这个小问题是可以轻松解决的：只要给在印第安纳的我接上一台连通静脉的喂食器，安排好它的程序，每当我——我的机器人身体——在加利福尼亚州收获了某种“食物”时，就往我的血管里注射营养成分即可（在加利福尼亚州的食物不需要是真的，只要我的远程机器人手臂做出取食的动作，就可以激活印第安纳家里那台连通静脉的喂食器）。

探索这些令人迷惑却又可行的远程呈现在“别处”的观念，会让人们逐渐意识到，随着远程呈现技术的发展，“首要”的定位会变得越来越不首要。实际上，人们大可以想象有一个传说中的“缸中的大脑”在布卢明顿控制着一个在加利福尼亚州巡回的机器人，全然相信自己是生活在西部的一个实体生物，而绝不相信自己是一个缸中的大脑。丹尼尔·丹尼特在他的哲学幻想小说《我在哪儿？》（*Where Am I?*）中，顺带探索了很多这样的观念。(注)

哪个视角才是真正属于我的？

关于要不要引入太多科幻小说式的场景来解释和论证我关于灵魂和意识的观念，我始终犹豫不决，因为这么做可能给人留下某种不好的印象：与我的视角在本质上紧密相连的是一个无药可救的科幻瘾君子不着边际的遐想；可我根本就不是一个迷恋科幻小说的人。尽管如此，我还是认为这样的例子通常有助于帮人们打破那些古老的、根深蒂固的成见。可是话又说回来，要想提醒人们意识到，日常的我们是如何司空见惯地把自己转移到虚拟世界中去，几乎也不需要讨论什么装在头顶的电视摄像头、远程控制的机器人或者连通静脉的喂食器。单是在起居室的窗边扶手椅上放松的休闲时光里阅读一本小说这个动作，就足以为这个现象提供一个绝佳的范例了。

当在读一本简·奥斯汀的小说时，我们眼中所见只不过是在一叠白色纸张上整齐排列成行的黑色墨迹而已，可是我们感觉到我们正在“看到”（我在这里应该使用引号吗？）的却是英国乡村的一栋家宅，在乡间小路上拉着一辆四轮马车的马群，衣着光鲜亮丽的贵妇和绅士肩并肩坐在马车车厢里愉快聊天时恰逢看到沿路破败小屋中走出一名穷苦老妪……我们太过沉浸于我们“看到”的事物了，以至于在某种重要且严肃的意义上，我们已经注意不到自己所在的房间或者窗外依稀可见的树

木，甚至连我们手里那些白色纸张上布满的黑色墨迹都注意不到了（反讽的是，即便如此，我们还是要依靠那些墨迹为我们带来我刚刚描述的视觉画面）。如果你不相信我，不妨想一想你在刚才那30多秒钟的时间里做过的事。你接触的是在白色纸张上散布的黑色墨迹，可是“看到”的却是在一间起居室的扶手椅上阅读简·奥斯汀小说的某个人，除此之外，你还“看到”了那座家宅、那条乡间小路、那辆马车、那对雍容华贵的夫妻和那位老妪……白色背景上的黑色花体字，经过恰当的安排之后，把我们在微秒之内传送到了一片无限遥远、早已逝去或者甚至从未存在过的时空当中。

所有这些都是为了重申一个观念，即我们能够同时人在好几个地方，在同一时刻欣赏不同的视角。你刚刚就做到了！你正坐在别的什么地方读这本书，可是就在不久之前，你又坐在某间起居室的扶手椅上读着一本简·奥斯汀的小说，而你同时还在乡间道路上的一辆马车上。至少有三个视角同时共存在你的大脑里。在这些观看者中，哪一个才是“真实”的呢？哪一个才是“真正的你”呢？这些问题需要回答吗？这些问题有答案吗？

我在哪儿？

我前几天开车出门，在等红灯的时候，停在了一个慢跑者旁边。她先是原地不动地小跑，然后等绿灯一亮，她就穿过马路，消失不见了。在那一瞬间，有种感觉仿佛她就是我。我之前从未见过她，之后可能也再见不到她了，但当时那种场景我却亲历过很多次。我以我自己的方式体验过那种经历，而且尽管我基本上对她一无所知，可我还是跟她共享了这种经验。可以肯定的是，我并没有透过她的眼睛来看世界。但还是让我们快速地向前跳跃一步，再来一次稍微有点儿傻兮兮的技术夸张。

设想每个人都在鼻梁上安装着一个微型的电视摄像头，而且每个人

还可以调节波段来接收地球上任意一台选定的电视摄像头传递的信号。如果有办法通过GPS定位锁定一个人（这似乎并没有那么玄乎），那么我只要把自己的眼镜设定为接收从那个慢跑者鼻子上的电视摄像头中发出的信号，说时迟，那时快！——我忽然间就在通过她的视角看世界了。当交通信号灯变色后，她跑走、消失，而我仍坐在我的车里，可我却能够跟着她跑走，看到她跑到的每一个地方，并能听到当她慢跑经过一段林荫小路时鸟儿的叽叽喳喳地鸣唱等。我还可以随时切换频道，通过我的女儿莫妮卡、我的儿子丹尼或者任何一个我选定的人的鼻尖摄像头来看这个世界。那么，我在哪儿？“你还在你原来的地方呀！”常识会如此叫唤说。可是这种说法太简单化、太模棱两可了。

是什么决定了“我在哪儿？”如果我们再次引入通过执行特定远程动作来吸收营养的观念，同时加上通过游戏手柄或者甚至某些大脑事件来控制远程动作的能力，那么事情就真的开始变得充满疑云、闪烁不定了。当然，控制行为机器人的电脑不可能恰好出现在与其无线相连的机器人所在之处。机器人可能正在月球上漫步，而它的计算机指挥系统则稳坐在地球上的实验室里。或者，一辆自动驾驶的汽车，比如斯坦利，可以横穿内华达沙漠，而它的计算机控制系统既可能安装在车内，也可能坐落在加利福尼亚州的一间实验室里，由无线信号相连。但是我们真的在意计算机在哪里吗？我们为什么要操心电脑的所在地呢？

在我们的感觉中，一台机器人就是它自己的身体所在。所以当我的大脑可以（使用上述的神奇眼镜）在上百个不同身体之间任意转换时——或者更糟，当它可以在同一时刻栖身于好几个不同的身体中并同时处理由它们接收的各种不同种类的输入（可能从一个身体接收视觉输入，从另一个身体接收听觉输入，而从第三个身体接收触觉输入）时，我在哪儿就变得相当说不清、道不明了。

身为另一个人的可变程度

让我们再一次把科幻小说式的场景抛在脑后，而只考虑日常的事件。我坐在一架正在降落的飞机上，身边人们各式各样的小段对话在不经意间飘入我的耳朵——评论印第安纳波利斯动物园有多棒，镇上新开的一家熟食店，诸如此类。每一小段对话都把我浮光掠影地带入另一个人的世界，让我体验到微乎其微的他人视角。我可能跟那个视角几乎毫无共鸣，可即便如此，我还是无比轻微地闯入了那个人的“私有”宇宙，而这次入侵虽然对于一个人类而言是绝对微不足道的，但是却比任何一只犬类动物闯入另一只犬类动物宇宙中的程度都要深。

而如果我跟另一个人就所有可以想象到的话题展开无尽绵长的谈话，包括最个人的感觉和最私密的坦白，那么我们两人世界的相互渗透就会强大到让我们的世界观开始融合。正如我在同身处Imlac室的斯科特·金通电话时可以跳跃到加利福尼亚州一样，每当有人呼喊出他们最热忱的希望和最揪心的恐惧时，我也可以通过言辞和语调随时跳到他人的脑中。

即使在一个完全非科技化的世界里，我们人类多多少少已经生活在其他人类的内部了。我们的大脑既然是表征性的通用机器，那么灵魂的相互渗透便不可避免。这也是“同情”一词的真正含义。

我有能力身为其他人，即便它只是生命行为的一种“简化”版本，即便它与那些人完整的力量与深度相差远甚。我有幸——至少我通常觉得自己是幸运的，虽然有时候我也会犹疑——总是拥有倒退并返回“单纯的我”的选项，因为在我的大脑里只居住了一个首要的自我。然而，如果在我大脑里有好几个强大的自我，彼此都在竞争首要的地位，那么“我”这个词的意义就真的要待价而沽了。

天真的视角往往就够好了

在我刚刚构想出的那个画面里，好几个自我在一个大脑内竞争一个首要的位置，这在你看来可能极其古怪，但事实上，这种发生在好几个“竞争自我”之间的内部冲突[注]是我们每个人都再熟悉不过的经验了。我们想要买一块糖，又想控制自己不买，我们对这种分裂感并不陌生。我们想驾车开完“最后20英里路”，又想在下一个休息站靠边停车，痛快地补上一觉，对于这二者之间的撕裂感，我们也深有体会。我们也许心里一边想着“再读完一段，我就去做晚饭”，一边又想着“我还是干脆把这章先看完吧”。这些对立的内在声音，哪一个才是真正的我？随着年龄的增长，我们学会了不要去问，也不要试图回答这样的问题。我们已经想当然地接受了这些小小的内部冲突，只把它们当作“人的境遇”的一部分。

如果你把左手放进一盆热水中，同时把右手放到一盆冰水里，一分钟之后，再把两只手迅速转移到一个温水缸里，你会发现自己的两只手——通常是你对于外部世界最信赖的侦察员和目击者——关于同一缸水产生了截然相反的感觉。面对这个矛盾，你最有可能给出的反应是耸耸肩，一笑置之，自己心想：“这是多么强烈的一种触觉上的错觉呀！”你不太可能在心里对自己说：“这次发生在我大脑内部的认知分裂，像一个楔子的锋利薄尖，揭露出在我的头脑中只存在一个自我这种日常信念的错觉性。”几乎每个人对于得出这个结论都是极为抗拒的，因为不管站在什么立场上，我们讲给自己听的那个简单版本的故事都足够好了。

这个情况有点儿类似牛顿力学。牛顿定律是极其可靠的，除非有物体以接近光速的速度向彼此靠近，它才会发生偏差[注]，给出大错特错的答案。可尽管如此，在绝大多数我们熟悉的场合里，甚至包括对前往月球或其他行星的宇宙飞行器轨迹的计算，都完全没有理由摒弃牛顿力学。这些飞行器的速度虽然相比于喷气式飞机已经提升了太多，可是相对于光速仍然微不足道，所以牛顿定律绝不可弃。

同理，关于我们大脑中有多少个灵魂的问题，我们为何要放弃常识

性的态度呢？我们很清楚，自己心里的答案就是一个。对此，我只能这样说：没错，答案很接近一个；但是当我们追问得更加深入一些时，我们便会看到，那个最初的精准估计值存在小小的偏差。不仅如此，我们甚至在日常的生活中也一直都能感觉到这种偏差——只不过我们更愿意把它们解释成无足轻重的错觉，或者干脆对它们视而不见。这种策略很管用，因为我们从来不会接近“光速”，所以那个天真的笼中鸟画面也不会崩塌。说得更直白一点，那些低分辨率、粗颗粒度的灵魂为了在我们的大脑中居住的机会而争得不可开交，但是它们从来没有真正对“第一号”的全面领导权构成任何实质性的威胁，所以那个天真的笼中鸟教条，“一个大脑，一个灵魂”，几乎从无争议。

锤头鲨觉得它在哪儿？

我们说一个单一的灵魂——比如说你自己的灵魂——被分装在好几个不同的大脑之中；对这一说法最有力的挑战，可能就是下面这个问题了：“好吧，让我们假设我以某种方式分散在很多的大脑里。那么我真正经历的是哪一个呢？我不可能同时既在这里又在那里吧！”但是我在本章中想要努力展现出来的，正是你确实能够在同一时间出现在两个地方，而且你甚至都注意不到任何蹊跷之处。你可以同时既身在布卢明顿，又身处斯坦福。你可以同时置身于唐纳山口（Donner Pass）滑雪小屋和中西部小镇上一家狗舍的游戏场地。你可以一边坐在自家起居室舒适的扶手椅上，一边坐在硌屁股的马车里，颠簸在19世纪的英国乡间的一条路上。

如果这些例子在你眼中太过不切实际，或者太科技化了，那么不妨想想智力低人一等的锤头鲨。这个可怜的生物，眼睛分别长在头的相反两侧，通常看到的都是两种完全不相干的场景。所以哪个场景才是这条鲨鱼真正看到的呢？说真的，它认为自己在哪里呢？当然，没有人会提出这个问题。我们就是接受了这一观念，即这种鲨鱼可以同时“差不

多”存在于那两个不同的世界之中，这主要是因为我们会劝自己说，不管那两个场景看上去有多么不同，它们仍然是作为鲨鱼生存环境的水下世界的连贯组成部分，所以到底在哪并不构成真正的问题。但这其实是在耍滑头，避重就轻。

为了让焦点更加集中，让我们对锤头鲨进行一点儿改造。我们设定一种生物的眼睛在接收一种场景（比如说在布卢明顿），而它的耳朵则在接受另一种不相关的场景（比如在加利福尼亚州）。同一个大脑在同时处理这些输入。我希望你可别把这说成是不可能的壮举！如果你想这么说的话，请先回忆一下，在你开车的时候，你还在跟其他的车辆、风景、指示牌和路标互动，同时还在跟手机另一端身在远方的朋友聊天（聊天中涉及的话题也许会把你如临其境地带到其他的地方去），而就在同时，你的脑袋里还在回响着最近听过的一段乐曲，你疲惫不堪的后背在困扰着你，你闻到了弥散在空气中的牛粪味道，你的肠胃还在朝你大喊：“我饿了！”你相当完美地应对着所有这些同时存在的世界——凭借同样的精神，没有什么能够阻挡一个人类的大脑同时应对来自斯坦福的声音和来自布卢明顿的视觉，这两个不相关的世界。正如锤头鲨的大脑也不会抗议说：“不要计算！”所以，“我不可能同时既在这里又在那里”这个观念，便烟消云散了。即便在我们每天的日常生活中，我们也一直都在同时身处这里和那里。

共振

但是你可能感觉到，我刚才的描述并没有回答最初提出的那个问题，关于你真正存在于那么多大脑中的哪一个里——不是在这儿，就是在那儿，意味着无论你和另一个人在情绪上有多么接近，他们的感觉永远是他们的，而你的感觉永远都是你的，二者永无相见之时。这又回到了本章开始的那个笼中鸟意象；这个意象必然不会停止屡屡抬起它那丑陋的头颅，不管我有多少次试图把它砍掉。可尽管如此，还是让我们再

试着采用另一种策略，继续对付这位美杜莎[注]。

如果我宣称，我有一部分生活在我的妹妹劳拉那里，而她也有一部分存在于我之中，那么即便如此，一个再明显不过的事实是，假如她现在碰巧正开车经过我们在圣何塞最喜欢的沙拉三明治店，并停下来点了一份沙拉三明治来吃，而我则正坐在远在印第安纳州布卢明顿的书房里拼命打字，根本品尝不到那个三明治。因此，我并不在那里，而是在这里！因此，我的意识也只在本地，而并非遍及全球，也没有四散开去！因此，故事到此为止！

可是事情并没有那么简单。我可能在一个小时之后接到劳拉的电话，听到了关于沙拉三明治的消息。当她生动地（或者甚至并不需要特别生动，反正我对那家的沙拉三明治已经那么熟悉了）描述起它时，我会记起那些小球的松脆质地，还有那美味的辣酱味道，口水也开始流下来。我对那沙拉三明治了如舌齿。虽然我的舌头并没有在爱抚那些小小的松脆的油炸小球，但我大脑中的某种东西已经从我所谓的“交感神经乐”（模仿“交感神经痛”而来）中产生了感官上的愉悦。尽管很微弱，而且延后了一个小时，可我还是分享了劳拉的愉悦。可即便这只是一次微弱的模仿而且并非同时，那又如何呢？即使我的快感只是她的一份低分辨率的拷贝，而且有时间上的错位，可是它仍旧是快感，仍旧是“关于”劳拉的快感，而并非关于我自己的快感。她的开心实实在在地传递给我。所以，虽然远隔两地、时间延迟、程度衰减，我还是披上了她的皮囊，而她也走进了我的内部。

这便是我想声明的一切——模糊性是存在的。发生在其他大脑中的某些事情，虽然不太清晰，但是仍然在“第一号”大脑的内部得到了复制，而两个大脑彼此在情绪上的距离越近，来回之间复制的东西也就越多，复制出来的副本也更加忠实。没人说这种复制行为是同时的、完美的或完整的——只是说，每个人部分地生活在另一个人的大脑之中，而如果带宽上调，不断提升，再提升，越来越大，那么他们就会越来越多

地生活在彼此内部——直到极限情况之下，他们之间界限分明的感觉将慢慢消解，如同双旋世界中对人（或者更像连体双旋世界中的对人）的两半。

现实是，我们并没有生活在一个像双旋世界那样的双胞世界之中，而且在我们生活的世界里，灵魂之间相对清晰的边界的存在也没有迫在眉睫地面临由带宽极高的脑际交流带来的威胁——大脑之间迅速而猛烈的信号交换，使得独立分别的身体不再能够判定独立分别的个人。现在的状况并非如此，而且我也想见不到它在可见的未来会成为现实。（但因为我并不是一名未来学家，所以也有可能说的不对。）

我的意思只不过是说，关于不同灵魂之间存在泾渭分明的界线这一神话的虚假性，我们所有人一直都有细微的体会。可是，把一个身体只与一个灵魂联系在一起的观念是如此便利而约定俗成，而把一个身体与一个灵魂视为精准的配对的观念又是如此诱人而根深蒂固，所以我们选择轻视或者完全忽视在日常生活中灵魂之间的相互渗透所具有的意味。

想想看，你能在多大的程度上让自己感受一位亲密好友的成功与失败，以及他们非常个人化的欣喜与苦恼。如果说，对于我妹妹那块沙拉三明治的间接享用在我而言是格外鲜活的，那么可想而知，当你常年孤身的朋友终于得遇佳人，开启了一段前景光明的姻缘时，或者当一位挣扎已久的演员朋友终于迎来一次幸运的突破，收获了媒体的盛赞时，你间接感受到的兴奋感将更加鲜活与强烈。或者反过来说，当你的密友突遭横祸时，你对于命运之不公所感到的怨愤又将多么有力。如果这都不算是你在自己的脑中过着他们的生活，那么该算什么呢？

话虽如此，我们在描述这类极其常见的现象时，使用的还是那些更加简单的、没那么有挑战性的说法，比如“他与她心存共鸣”、“她真是个有同情心的女子”、“我理解你所经历的一切”、“我与你感同身受”、“看到她的遭遇令我心痛”或者“别再跟我说了——我受不了！”这些标准的表达方式，虽然确实反映出某个人部分地存在于另一个人的内

部，却极少甚或从未有人把它们的字面含义当真，推断出我们的灵魂真的在彼此渗透并模糊在一起。这个观念太过棘手，而且可能甚至太吓人了，叫我们无法应对，所以我们还是更愿意坚持认为，我们就像相距甚远的星系一样，彼此之间并没有真正的重叠。深信不疑地接受灵魂的笼中鸟隐喻已经成为我们一生中根深蒂固的习惯，要想打破这样一种深入骨髓的习惯是相当艰难的。

我不是任何别人吗？或者说，我是每一个他者？

从根本上说，笼中鸟意象中隐含的本义是，不同的人就像同一条线上独立的点，这些点的直径为零，所以无论如何都不可能相互重叠。确实，如果我们拿初等代数中所谓的“实数线”（real line）来做一隐喻，那么笼中鸟隐喻就会给每个人分配一个“序列数”——一个无限小数，唯一地决定着身为那个人“是什么样子”。由此观点出发，则不管我们觉得我们有多么相似，也不管我们在人生中分享了多少共同的经历，即便我们是全等或连体双胞胎，你和我还是在出生的时候就被分配了不同的序列数，因此在那条线上占据了两个不同的零宽度点，这是板上钉钉的事情。你是你，我是我，不管我们距离多近，都毫无重叠。我不可能知道身为你是什么样子，反之亦然。

相反的论点会主张，每个人都均匀分布在整条实数线上，因而所有的个体都是同样的一个人！只有一个人存在。虽然这个极端的观点没有得到普遍的支持，可依然有其在现代的拥护者，比如最近出版了《我是你》（*I Am You*）一书的哲学家丹尼尔·科拉克（Daniel Kolak）。这个观点在我看来几乎没有任何道理可言，简直跟泛心论一样；后者宣称所有事物都是有意识的[注]——石头、餐桌、野餐、电子、彩虹、一滴水、瀑布、摩天楼、炼油厂、广告牌、限速标志、交通罚单、县监狱、田径运动会、选举舞弊、机场大门、春季甩卖、肥皂剧停播、玛丽莲·梦露

的照片，我还可以没完没了地列举下去，直到令人作呕。

本书的观点介于这两种极端的观点之间，没有把个体描画成点状的无限小数序列数，而是沿着那条线四处散布的范围相当集中而又模糊不清的区间。在这些区间，有一些重叠得很严重，但大多数都极少重叠，甚至完全没有重叠。毕竟，两片一英尺宽的模糊区间如果相距100英里远，那么它们的重叠度显然是零。而两片一英尺宽的模糊区间如果相距只有半英尺，那么二者之间肯定有大量的重叠。在这样的两个人之间，将不会有一道不可跨越的存在性裂痕。相反，他们每个人都延展到另一个人之中，部分地生活在另一个人的内部。

国魂的相互渗透

在本章更早的部分，我简略地勾勒过一个把自我类比于在很多其他国家设立大使馆的国家的意象。下面，我想追问一个相近的议题。但是首先，我要从国家为何物开始，重拟一个充分简化的概念，然后在此基础上展开。所以，让我们先来看看“一个国家，一个民族”这个口号。这个口号中包含的意思是，每个民族（一个包括了历史、传统、语言、神话、文学、音乐、艺术、宗教等等在内的精神性文化概念）总是和某个国家（一个包括了海洋、湖泊、河流、山脉、峡谷、草原、矿藏、城市、公路、明确的法定边境等等在内的物理性的地理概念）精确而规整地结合在一起。

如果我们真心相信人类自我的笼中鸟隐喻有这样一种严格的地理类比物，那么我们就会产生一种奇妙的信念，即所有来自某个特定地理区域内部的个体总是拥有相同的文化认同。像“美国人在巴黎”这样的短语在我们这里是说不通的，因为法国国籍与被称为“法国”的那片物理地区的边界保持严格的一致。在法国不可能有美国人，而在美国也不可能有法国人！当然，这个类比性的概念要适用于所有的国家和民族。这显然

是荒诞不经的谬论。移民和旅行都是普遍的现象，而它们正在不停地把不同的国家与民族搅和在一起。

可这当然并非意味着民族或者国家是根本不存在的东西。尽管两个概念都具有巨大的模糊性，可它们依然是有用的。例如，让我们来想一想意大利。西北部的瓦莱达奥斯塔大区（Valle d'Aosta）主要通行的是法语，而东北部的南蒂罗尔（Südtirol）主要通行的是德语。不仅如此，在米兰以北的国界之外，瑞士的提契诺州（Ticino）却是以说意大利语为主。所以，意大利的国家与意大利的民族之间到底有什么关系？至少可以说，它们的关系不是精确而分明的——可是意大利和意大利人仍然是对我们有用的说法。只不过，我们知道这两种概念是模糊的。而意大利的情况也是每个国家的情况。我们知道，每种国籍都是一个模糊的、四散的现象，以某个单一的地理区域为中心，却不限于此，而我们也都完全习惯了这种概念，一点儿也不觉得它自相矛盾或者混乱不清。

所以，让我们充分利用这种对于地区和民族之间关系的舒适感，努力在身体和灵魂之间的关系上获取一种更加成熟而深刻的把握。想想中国，在过去的几个世纪中，有上百万人移民国外。中国是干脆把这些人都忘记了吗？中国有没把他们视为逃亡者而从它的集体记忆中删除了呢？完全没有。在中国境内，还留存了一种对于“海外华人”的强烈情感。这些人虽然远在天涯，却备受珍视，中国呼唤他们“回家”，至少是短暂的探望，而当他们还乡之时，会像失散已久的亲人一样（当然，他们本来就是），受到热情的欢迎。如此，中国的这条海外分支在中国国内被认为是中国不可缺少的一部分。它是中国性的“光晕”，远远地超出了国土的物理疆界。

当然，不仅是中国，每个国家都有这样一圈光晕，这圈光晕在地球上每个其他的国家里闪烁着光芒，时而耀眼，时而黯淡。如果在国家的层面上也有一种与人类的死亡对应之物，那么当一个民族的“身体”被（如巨大陨石坠落等大灾难）消灭之后，因为他们在国土的物理疆界之

外还存在闪亮的光晕，所以这个民族仍未消亡，至少是部分地幸存了下来。

这个画面虽然有点可怕，但是一点儿也不会给我们留下违反直觉的印象，因为我们都理解，物理的国土，不管它在颂歌与传说中收获了多少爱意，对于一个民族的存活而言，其实并不是不可或缺的。地理位置只不过是一套古老的基因与迷因——肤色、体型、发色、传统、词汇、谚语、舞蹈、神话、服饰、食谱等等——的传统滋生地，而只要在这些基因和文化基因的携带者中，身在国外而逃过灾难的数量达到了临界的标准，那么所有一切丰富性都可以继续存在下去，并在别的地方蓬勃活跃起来，而那个如今已逝的物理地点还可以继续在歌曲与传说中得到传诵。

虽然从来没有哪个国家曾遭到过物理上的整体毁灭，但过去还是多少发生过与此类似的事情。我想到的是18世纪和19世纪波兰的左邻右里对波兰国土的吞食——所谓的“瓜分波兰”（partitions of Poland）。波兰人虽然在物理上已无家可归，可他们还是坚韧地存活了下来。波兰的国家依然在——波兰人（naród polski）——充满生机，活力不减，只不过他们的每一寸领土都被剥夺了。实际上，波兰国歌的第一句歌词便是在歌颂这幸存：“只要我们一息尚存，波兰就决不灭亡！”无独有偶，原初的犹太人自《圣经》时代开始就已告别了他们文化的摇篮而流散各地，可是犹太人在大范围的流散中继续存活下来，延续了他们的传统、语言和信仰。

光晕、余晖、日冕

人死之后，留在人间的是一组余晖，有些更耀眼，有些更黯淡，在所有与他们亲近之人的大脑中闪烁。等到那些人过世后，余晖会变得格外昏暗。而当最外面的一层也被遗忘后，余晖会更加虚弱，直到再过一

段时间后，一丝光亮也不剩了。

我刚刚描绘的这一缓慢消亡的过程，虽然前景终归幽暗，可相比于标准的观点却减少了一点点阴郁之感。因为身体的死亡太过清晰分明，也太过戏剧化，而我们又总愿意死抱着笼中鸟的观点不放，所以死亡在我们心中的印象是即时而绝对的，像断头台的刀片一样锋利。我们本能地相信，光也在同时熄灭了。依我所见，人类灵魂并非如此，因为一个人类生命的实质——真的跟一只蚊子、一条蛇、一只鸟或一头猪的实质不一样——分布在很多大脑之间。要经过几代人的时间，一个灵魂才会消逝，那闪光方始停息，那余烬终得燃尽。虽然“尘归尘，土归土”可能是最终的真相，但它所描述的转变并不一定如我们以为的那样骤然。

因此，在我看来，举办葬礼或追悼会的目的，虽然极少明言，却是在本能地想要把与死者最亲密的众人重聚起来，在所有人心中最后一次集体点燃那特别的生命火焰。那火焰代表的是他们所爱之人的实质，从彼此的在场中直接或间接地汲取光亮，并与尚存世间的大脑分享那个人在场的感觉，由此在最大的可能限度上巩固那些次要的个人胚芽；这些次要的个人胚芽仍在所有那些彼此不同的大脑中闪烁发光。虽然首要的大脑已经暗灭，可是在那些聚起来回忆并重新唤起逝者精神的人中，尚有一轮集体的日冕在熠熠生辉。这就是人类之爱的意义。因此，“爱”这个字眼不能与“我”这个词分开讨论；一个人的符号在你的内部扎根越深，爱就越深，而那道留在身后的光也就更加明亮。

---

1. 一个灵魂散布在很多大脑之间的想法在我的脑海中形成的是一幅固态物理学的画面，而我博士期间读的正是这个专业。固体就是晶体，意味着一个在空间内周期性排布的原子网格，就像果园里的树，但其组织形式是三维的，而不是二维的。在某些（绝缘的）固体里，电子“盘旋”在每个原子核周围，与原子核紧紧地绑在一起，不能脱离。它们就像绕着果园里一棵大树盘旋的蝴蝶一样，只不过从来不敢冒险跑到相邻的另一棵树那里去。相反，在作为绝佳导体的金属中，电子不会怯懦地守住一棵树，而是英勇地浮动在整个网格之间。这就是金属为什么导电性那么好的原因。事实上，一只蝴蝶在大树之间漫无目的地乱飞的图像并不适合用来描绘一个金属中的电子，后者更像是同时分布在

整个晶体中的一种强度模式——某些区域更强，某些区域较弱，不断变化。一个电子也许更应该被比作一整群橘色的蝴蝶，而另一个电子就是一群红色蝴蝶，或者一群蓝色蝴蝶，每一群都散布在整个果园里，相互混在一起。简言之，金属中的电子绝不是紧密连接的点；它们是浮动的模式，居无定所。但是让我们不要忘了所有这些意象的意义所在，那就是为想象人类灵魂的实质提供有益的方式。如果我们把晶体网格中的每棵树（或原子核）与一个特定的人类大脑对应起来，那么在束缚紧密的模型（相当于笼中鸟隐喻）中，每个大脑都拥有一个独特的灵魂，表现为一小群怯懦的蝴蝶，紧紧守着那棵树，寸步不离。相反，如果我们把它想象成金属，那么那片云就散布到整个网格之间——也就是说，被所有的树（或原子核）平等地共享。没有哪棵树是有优先特权的。那么，在这幅景象中，（很接近科拉克在《我是你》中的观点）每个人类灵魂都浮动在所有人类大脑之间，而它的同一性并不是由它的位置决定的，而是取决于它形成的波动不定的整体模式。这些都是极端状况，但我们也完全可以想象一种折中的情况，很多群有各自定位的蝴蝶在一棵树附近浮动，但是不仅限于那棵树。这样一来，一群红色的蝴蝶也许以树A为中心环绕，但是也可以蔓延到邻近的10多棵树那里，而一群蓝色的蝴蝶也许大概以树B为中心，一群黄色的蝴蝶则以树C为中心，如此等等。每棵树都是一群蝴蝶的中心，而每群蝴蝶也都只有一棵主树，但是这些蝴蝶群亲密地互相渗透，以至于很难分清哪一群“属于”那棵树，反之亦然。这个诡异而超现实的故事起始于固态物理学而终于果园里互相渗透的彩色蝴蝶群，这是我能够为描述人类灵魂如何分散大脑之间所绘制的最清晰的画面了。

2. This classic piece can be found in [Dennett 1978] and in [Hofstadter and Dennett].

3. Chapter 13 of [Dennett 1991] gives a careful discussion of multiple personality disorder. See also [Thigpen and Cleckley], from which a famous movie was made. See also [Minsky 1986] and Chapter 33 of [Hofstadter 1985] for views of a normal self as containing many competing subselves.

4. See [Hoffmann] for a discussion of the subtle relationship between relativistic and Newtonian physics.

5. 美杜莎（Medusa）是古希腊神话中的蛇发女妖，任何人看到她的脸都会变成石头，在传说中被视为极其丑陋的象征。此外，她还是戈耳工三姐妹之一，她的两个姐姐都有不死之身，只有美杜莎是凡人，可以被杀死。作者在此用美杜莎类比“笼中鸟”隐喻，也暗含这个隐喻可以被攻破之意。——译者注

6. See [Rucker] for a positive view of panpsychism.

第19章
意识=思考

那么，在我的环路故事里，意识在哪儿？

从本书第一页开始，我对于下面几个关键术语的使用基本上是可以互相替换的："自我""灵魂""我""内部之光"和"意识"。于我而言，这些都是同一种现象的不同名称。在其他人看来，它们看上去可能并不是指同一件事，但是在我看来并非如此。这就好像满足4n+1形式的质数和等于两个平方数之和的质数——表面上看，它们似乎在描述两种完全不同的东西，可是加以更仔细的分析就会发现，它们的所指原来是完全相同的东西。

依照我看待事物的方式，则所有这些现象都有深浅不同的灰度，而不管它们其中之一在某个（自然或人造的）特定生命中深浅如何，其他的现象都将具有同样的灰度。所以我才会感觉到，当我在谈论"我"的时候，其实一直也在谈论意识。话虽如此，我深知会有人抗议说，虽然我也许讨论了个人同一性的议题，可能也触及了"我"和"自我"的概念，但是关于意识的那个远为深邃而神秘的谜题，我甚至连边儿都没沾上。他们会怀疑地质问我："那么，在你的怪圈学说里，什么才是经验？大脑中的怪圈如何告诉我们，活着、闻到金银花的味道、观看日落、聆听雨滴屋檐的声音，是一种什么样的感觉？这才是意识的重点！这与你那个怪圈观念有什么关系呢？"

面对这些如铁板一块的怀疑论者提出的这类问题，我不确定能否给出令他们满意的答案，因为他们肯定会觉得我说的话既幼稚简单，又闪

烁其词。不管怎么说，我还是把我的答案放在这里，芟除芜杂，直陈精义：意识是闹壳内部的符号之舞。或者，甚至还可以更加简洁扼要地说，意识就是思考。正如笛卡儿所言："我思故我在。"

遗憾的是，我怀疑即使在与我最相投和的读者看来，这个答案也太过精简了，所以我打算想办法把它说得更明白一点儿。在绝大多数时候，我们大脑中任意一个给定的符号都处于休眠状态，就像一本书了无生气地躺在一座巨大图书馆偏僻的藏书架上。有时候，某个事件会触发从书架上抽出这本书的动作，然后这本书会被翻开，而它的书页则在某个读者面前活了过来。在一个人类的大脑中，感知到的外部事件也在以类似的方式不断高度选择性地触发把符号从休眠中提取出来的动作，让这些符号在各种各样意想不到的、前所未有的设定中苏醒过来。这种大脑内部的符号之舞，便是意识。（它也是思考。）注意，我说的是"符号"，而不是"神经元"。舞蹈必须在前一个层面上被感知到，才能构成意识。以上就是更加清楚一点儿的解释版本了。

走进怀疑论者

"那么，谁来解读这些符号和它们的组态呢？"有些怀疑论者会问，"是谁感觉到这些符号'活了过来'？与取书的读者相对应的主体在哪里？"

我猜，这些怀疑论者会争辩说，符号之舞就其本身而言只不过是物质材料的运动，没有被任何人感觉到，因而不管我如何主张，这支舞蹈还是无法构成意义。这些怀疑论者想让我点名指出我们每个人都拥有的那种关于自己思想与感知的主观自觉性（awareness）的特殊所在。可我却感觉，他们的这种希望很令人费解，因为它启用的"自觉"一词在我看来不过是"意识"的另一个同义词而已，只是在另一个层面上再次提出了同一个问题。换句话说，人们在追问被激活符号组态的"读者"时，可

能已经接受了大量符号在大脑中受到触发这一观念，但是他们拒绝把那种内部的翻滚蒸腾称为“意识”，因为他们现在想让符号本身被感知到。[注]如果我在这个节点重提闹壳隐喻，并指出闹壳里的小磁交弹球之舞也能构成意识，可能会让这些人感到格外反感。他们会反对说，那只不过是一大群小小的玻璃球在一张台球桌上来回碰撞而已，显然是空洞而缺乏意识的。他们想要的比这更多。

这些怀疑者实质上是在把问题明升暗降——反映外部事件的符号层面的大脑活动（或者小磁交弹球层面的大脑活动）就是意识，这个观念不能令他们满足；他们现在坚持要求的是，当意识发生时，大脑活动的内部事件也要被感知到。这就带来了无限回归的风险，非但不能离答案更近一步，反而距解开意识之谜越行越远。

不管怎么说，我可以给这些人一些认可——我同意，符号的活动本身就是一个人类大脑重要而不可或缺的注意焦点（但是我要迅速追加一句补充，对于鸡、青蛙或蝴蝶并非如此，对于狗也基本上不成立）。成熟的人类大脑在持续不断地努力降低他们感知到的东西的复杂性，这就意味着他们在持续不断地努力让由很多同时激活的符号组成的陌生而复杂的模式仅仅触发一个熟悉的先在符号[注]（或者很小的一个符号集合）。事实上，这正是人类大脑的主要使命——确切指出一个复杂情景中的要紧之物，从初始的一团感觉与观念的大杂烩里提炼出这个情景真正关乎之事。直击要点。然而，对于斑点狗而言，这个要点是无关紧要的，而对于斑点狗摇动的尾巴上的跳蚤而言，这个要点当然更无关紧要。

我担心以上所有这些听起来都有点晦涩难解，所以下面用一个典型的例子来进一步阐明。

符号触发更多的符号

一位名叫妮可的学生有可能来我的研究小组读博士，于是她前来参观了一天，衡量一下来此深造的可能性。我和我的研究生跟她进行了好几个小时的交流，先是在我的研究中心谈话，然后还在一家中餐馆共进晚餐。我们一致认为，她拥有令人愉快的活跃思维，她的思想也正好跟我们同在一个频段，而且很明显的是，我们的热情也并非是单方面的。然后，不用说，我们对于她在下个秋天加入我们都抱有很大的希望。妮可在回家以后，给我发了一封电子邮件，说她还在为我们的观念感到兴奋不已，那些观念鲜活地回响在她的脑海中，绕梁不绝。我回了一封鼓励性的邮件，然后在接下来的几周里，就杳无音信了。后来，我终于又给她发了一封邮件，告诉她我们都很盼望她明年能过来。几天之后，她回了一封语气多少有点儿僵硬的简信，说她很抱歉，可是她已经决定去另一所大学读研究生了。“但我希望我们在未来还有交流的机会。”她在信末有礼貌地补充说。

好吧，这段小小的插曲于我而言可是完全新鲜的经历。妮可是一个独一无二的个体，我们和她充满活力的谈话都是那么独具一格，而整件事在我的大脑中激活的复合符号组态，确实是史无前例的。可是从另一个层面上来说，却并非如此。

其实我只要“随便想想”就会发现，在我积累了好几十年的情景记忆里，这个情景有过大量的先例。事实上，根本不需要我付出什么努力，不少多年不曾重温的过往记忆便会涌现上来，比如在差不多30年前，一位很有希望成为我们同事的候选人看上去对于这份工作特别感兴趣，可后来让我们大出所料的是，他竟然拒绝了我们给出的超乎慷慨的邀请。过了几年之后，我有一名极为聪明的研究生对于跟随我前往加利福尼亚州度过学术休假表示兴奋不已，可是他随后也改变了主意，而且很快在我的视野中完全消失了，再也没有消息传来。另外，还有一段令人伤感的记忆。我深深地迷恋上了一位来自远方的妙龄女子，而她向我发出的信号最初也似乎充满了撩拨人心的应许，可是后来，她莫名其妙地冷淡了一些，又过了差不多一个星期，她才告诉我，她其实正处于跟另外一

个人的关系当中（事实上，令我郁闷的是，那个事件远非仅仅发生过一次）。

所以，一件接着一件，所有这些积满尘灰的旧“书”，都在当前情景的唤醒下，从休眠的书架上被抽取了出来，因为当我们在抽象的层面上感知这一“史无前例”的场景，剥去它的外壳而提炼出它的内核(注)，它便直指向存储在我的“图书馆”架上的某些过去的长篇故事，而它们一本接着一本地被抽取出来，并被放在了激活的白炽光下。这些古老的长篇故事，在很久很久以前被干净整洁的心灵打包起来，悠闲地坐在我大脑的书架上，等待着被“同样的事情”触发；当然，前提是这些事情在改头换面后真的会再次发生。而我不得不难过地说，它真的发生了！

记忆触发触发记忆的记忆，当所有这类活动展开一段时间之后，有些东西会慢慢地稳定成型——借用化学的术语说，就是某种“沉淀”。在我们的案例中，最终归结出来的一个词就是：“被涮”。没错，我感觉自己被涮了。我的研究小组也被涮了。

这是对复杂性多么显著的一次约简啊！我们从一次持续了数个小时并辗转了两个场地的会面开始，这次会面涉及了许多人、数千字的交谈、数不清的视觉印象以及几封事后跟进的邮件往来，但最终，整件事只过滤出了（或者我是不是更应该说是“栽在了”？）一个令人失望的双字词“被涮”。可以肯定的是，这并不是我从这段故事中留存下来的唯一观念，但是“被涮”成为一个主导性的心灵范畴，而妮可的来访将跟这一范畴永远地联系在一起。当然，妮可的故事本身会被干净地打包起来，存储在我的情景记忆书架上，留待在未来的某一天可能被我的这个“我”在此提取出来，谁知道那会在什么时候，什么地方。

认知的中央环路

这种美妙的流动型抽象认知与记忆提取所依赖的机制，跟上述怀疑论者鼓噪之物多少有一点儿相似——这是一种对于内在符号模式的感知，而不是对外在事件的察觉。某个人似乎正在看着被激活的符号的组态并感知到它们的实质，由此触发了其他休眠符号的提取（如我们刚才所见，可能是很大的结构——例如，存储了整段长篇浪漫故事的记忆包裹），如此交替往复，引发了符号活动的活力循环——完全是临场发挥而又流畅无比的符号之舞。

这种形成符号触发循环的阶段，乍一看，可能跟你从海量的视觉输入中认出一棵木兰花树是完全不一样的，因为后者涉及的是对外部场景的处理，相比之下，我在此则是在看着我自己被激活的符号的舞蹈，并试图精确地描述这舞蹈的实质，而不涉及某些外部场景。但是我要提出的主张是，二者间的裂缝远没有人们第一印象中那么遥远。

我的大脑（还有你的，亲爱的读者）在不断地想要贴上标签、划分范畴和寻找类比与先例——换言之，就是想要加以简化而又不让实质溜走。(注)它不知疲倦地执行这项活动，不仅是在回应新的感官输入，而且也在回应它自己的内在之舞，而这两种情况之间其实并没有什么差别，因为一旦感官输入越过视网膜、鼓膜或皮肤，进入内部的领域，感知就已经完全是内部的事了。

简言之，应该可以让怀疑论者感到满意的是，真的存在一种符号活动的感知器——但是让他们不满意的是，这种“感知器”本身只不过是更进一步的符号活动。没有什么会有魔法发生的特殊“意识之所”(注)，没有什么除了差不多同样的东西以外的东西，也没有什么与舞蹈的符号发生联系的某个地点……好吧，那是什么？什么能让怀疑论者感到满意？如果“意识之所”原来只不过是大脑的一个物理部分，这就能让他们感到满意吗？他们还是会抗议说，如果那就是我所宣称的意识，那么它就只是毫无生气的物理活动而已，跟那些小磁交弹球在毫无生机的闹壳场地里像无头苍蝇一样乱冲乱撞没什么区别，也没好到哪里去，更是同意识

毫无关系！

我想，到了这一步，有必要把我内心各种各样的怀疑声音融合成一个纸面角色（希望不是一只纸老虎），再让这个人物跟另一个实质上代表本书观念的角色展开一场延长的辩论了。我把这本书的代言人称作“怪圈641号”，而把怀疑论者的代言人称作“怪圈642号”。

有些读者可能会认为，我对于此事的处理充满了不公正的偏见，因为我不但给我自己（或者不如说是我的代言人）贴上了“怪圈”的标签，还把这个标签强加给我值得尊敬的对手身上。这样一来，仿佛游戏尚未开始，就已经结束了。但是，标签只是标签。在这段对话里，重要的是人物们说了什么，而不是我怎么称呼他们。所以，如果你更愿意把怪圈641号和642号分别命名为“内在之光7号”和“内在之光8号”，或者甚至是“苏格拉底”和“柏拉图”，我都是没问题的。

下面，闲话少说，让我们调转频道，让我们的两个怪圈（或者内在之光）开始亲切的辩论吧。哎哟！我猜，我好像又在这里多唠叨了几句，看来我们已经很遗憾地错过了两位朋友开场时的一些机智的交锋。噢，好吧，生活就是这样。但愿你和我现在跳转频道进去，还跟得上。让我们试试吧……

---

1. See the careful debunking in [Dennett 1991] of what its author terms the “Cartesian Theater”.
2. 这句话对于制作一张索引表的噩梦般的工作格外适用。只有当一个人把几周的时间花费在一张精细的索引表上时，这个人才能真正理解这项任务有多么残酷（和荒谬）。
3. See [Sander], [Kahneman and Miller], [Kanerva], [Schank], [Boden], and [Gentner *et al.*] for discussions of the analogy-based mechanisms of memory retrieval, which underlie all human cognition.
4. See [Hofstadter 2001], [Sander], and [Hofstadter and FARG]. To figure out how to give a computer the rudiments of this ability has been the Holy Grail of my research group for three decades now.

5. See [Dennett 1991].

第20章
一场谦恭有礼的辩难

剧中角色：

怪圈**641**号：《我是个怪圈》观念的信仰者

怪圈**642**号：《我是个怪圈》观念的怀疑者

* * *

642 可怕啊，真是太可怕了。事实上，你勾勒出的灵魂画像不仅仅可怕，它完全是空的。空虚。那里根本就没有任何精神性的东西。只有物理活动，仅此而已。

641 你还期待有别的什么吗？你还能期待有别的什么呢？除非你是个二元论者，也就是说，你认为灵魂是鬼魂般的、非物理性的东西，不属于物理世界，却能在物理世界中作威作福。

642 不，我不信那一套。只不过，总得有某些极为特殊的东西，对物理世界中那些生命负责，那些生命具有精神、心灵、感觉和感知——而那种特殊的东西可以解释我们内在之光、我们的自觉、我们的意识。

641 我对此深表赞同。要想解释如此难以捉摸的现象，必然需要某种特别的东西。用物理零件建造出一个灵魂，看似是离谱的要求。但是不要忘了，在我的观点中，意识是一种极其罕见的复杂组织的物理模式，而不仅仅是传统的物理活动。不是链条的摇摆，不是石头扑通一声落入池塘，不是瀑布溅起水花，不是飓风的涡旋，不是抽水马桶的上

水，不是房间里温度的自动调节，不是国际象棋程序中的电子流动，不是追求卵子的精子在摇摆尾巴，也不是一只饥饿的蚊子大脑内的神经放电……但是随着这份清单不断推进，我们也越来越接近目标了[注]。当我们沿着这个等级结构向上攀登时，一道“内在之光”便开始亮了起来。虽然这道光在这份清单的终点处仍然昏暗得令人难以置信，可是如果我们进一步延长这份清单，向上扫过蜜蜂、金鱼、兔子、狗和婴幼儿的大脑，它就会变得明亮很多。当我们成为人类青少年和成年人时，这道光就非常明亮了，而且它还会持续明亮几十年。没错，我们所知的属于我们自己的意识只不过是人类大脑中的物理活动，在这个世界上活跃了数个年头而已。

642 不对，在你的图景里，意识的实质是缺失的。你描绘了一套复杂的大脑活动，涉及符号之间的相互触发，而我也愿意相信，大脑内部确实发生了类似的事情。可是那并不是故事的全部，因为我在这个故事里没有任何立足之地。这里面没有给一个我留下任何容身之处。你提出的是大量无意识粒子的来回冲撞，或者也许是由粒子组成的大片的活动之云——但是如果宇宙仅此而已，那么就没有我没有你，也没有任何一种视角可言了。这是生命进化前地球的样子——数百万年的日出日落，云聚云散，风四处吹，雷暴冲荡山谷，巨石滚落高山而形成沟壑，水入河床而雕出深深的峡谷，波浪拍打沙滩，潮来潮退，火山喷出滚热的熔岩之海，山脉从平原中崩裂而起，大陆漂移、分裂，如此等等。虽然这都是壮美的景观，但是没有内在生命，没有思维，没有内在之光，没有我——没有人来欣赏这伟大的景象。

641 我跟你有同感，一个仅仅由物理现象构成的宇宙是荒寂的，但是某些种类的物理系统，可以反映出在他们外部的事物，还能够依据它们的感知来发出行动。这才只是初露端倪而已。当感知变得足够精密复杂之后，它可以引出在仅仅以原始方式进行感知的系统中找不到对应物的现象。所谓“原始”感知系统，我指的是比如恒温器、膝盖、精子、蝌蚪之类的实体。这些实体太过初级，还配不上“意识”这个说法，可是，

当发生感知的系统拥有一套真正丰富而流动的可延展性符号集合时，一个“我”便会应运而生，就像怪圈在《数学原理》寸草不生的堡垒中拔地而起一样。

642 感知？！谁在感知？没人！你的宇宙仍然是一个虚空的系统，只有物理实体和它们错综盘绕、纠缠不清的运动——星系、恒星、行星、风、石头、水、滑坡、涟漪、声波、火、放射活动等等。甚至还有蛋白质、RNA和DNA。甚至包括你心爱的反馈环——跟踪导弹、恒温调节器、抽水马桶、视频反馈、多米诺链、布满成群的微型磁力小球的台球桌。但是在这片荒凉的场景中，缺少某种关键的东西，那就是“我”。我在一个特定的地方。我在这里（hereness）！如果一个世界里只有成千上万个水桶中的水和浮球，或者只有亿万个不同的多米诺链，那么靠什么才能显出一个这里呢？那里根本就没有这里。

641 这个问题会给你造成困扰，我真心理解这一点：任何一个思考这个问题的人都会受到这样的困扰。我的回答是这样的：你刚刚已经无比生动地提及了各种物理实体，在由多种多样物理实体组成的广袤宇宙中，存在某些罕见的局部活动的地点，在那里可以发现一类特殊的抽象涡旋模式。那些罕见的地方——至少就我们目前为止所遇见的那些而言——就是人类的大脑，而“我”们正局限于那些场所之内。这些场所在广大的宇宙中很难发现；它们数量既少，相隔又远。但凡这种特殊的、罕见的物理现象发生，一个我和一个这里就出现了。

642 你用到的“抽象涡旋模式”这个短语，让我想到了一种物理涡旋，比如飓风、旋涡或者螺旋星云[注]——但我想这些东西对你而言都不够抽象。

641 是的，他们确实不够抽象。旋涡和飓风只是旋转的涡流——是陀螺和回转仪的流体近亲。为了制造一个“我”，你需要意义，而为了制造意义，你需要感知和范畴——事实上，你需要一个不断自我增长的范

畴库，不停地扩充、生长、扩张。这种东西在你提到的物理涡流中是找不到的。正是因为如此，我们才需要为“我”找到一个更加贴切的隐喻，那就是哥德尔在表面寸草不生的*PM*宇宙中发现的自我指涉的公式结构。他的公式跟人类的“我”一样，结构是极端繁复而精密的，而且可谓罕见。“平凡”的*PM*公式，比如“0+0=0”或者陈述了每个整数最多只能写成四个平方数之和(注)的公式，可以类比于死气沉沉、不含“我”的物理实体，比如沙砾或者保龄球。那种简单的公式没有包在外面的高层级意义，可是哥德尔的特殊符号串有。从关于数字的平凡论断中构建起哥德尔式怪圈的复杂性，需要耗费大量的数论装置（machinery），同样，从最简单的反馈环中构建起大脑中怪圈的复杂性，也需要漫长的生物进化历程。

642　　我暂且接受你的说法，假定在宇宙中四处漂浮着很多抽象的“怪圈”，它们在数十亿年的进化过程中以某种方式联合在一起——这些怪圈居住在大脑的内部，有点像音频反馈环在音乐厅中安身。它们可以要多复杂有多复杂，随你高兴；我压根儿就不关心它们的物理活动复杂性。可是那个关键性的问题仍然没有解决：那些怪圈中的一个如何才能成为<u>我</u>？哪一个怪圈才是我？你无法回答。

641　我能回答，虽然你可能不喜欢我的答案。它们之中成为<u>你</u>的那一个之所以成为你，是因为它居于一个特定的大脑中，而那个大脑经历了所有让你成为你的经历。

642 这就是句循环论证的废话！

641　非也，非也。这个观念有点微妙，其中的关键在于，被你称作“我”的东西是一种<u>结果</u>，而不是一个起点。你是通过一种没有事先计划的方式聚合起来的，慢慢地成为存在，而不是一下子就出现了。最开始，当后来成为你灵魂之所的大脑刚刚形成的时候，那里面还没有你。但是那个大脑缓慢地生长，它的经历也在缓慢地累积。随着越来越多的

事情发生、被它收录下来并完成了内化，到了这一过程中的某一点，它开始模仿自己浸淫其中的文化和语言的习俗，进而试探性地用“我”来称呼自己（尽管这个词的所指之物仍然十分模糊）。大概就是在这个时候，它注意到它自己身在某处——不出意外，它就是身在某个特定大脑的所在之处！但是在这个节点上，它对于自己的大脑还是一无所知。它只知道它大脑的容器，也就是一个特定的身体。可即便它对于自己的大脑一无所知，那个初生的“我”还是会忠诚地追随着它的大脑，如影随形。

642 你没有正面回答我的问题。我问的是，在一个由无法分辨的物理结构组成的世界里，怎么才能显出我来。

641 好吧，我来直接回答这个问题。在你看来，所有容纳怪圈的大脑看起来都跟成千上万台四散的缝纫机没什么差别，大同小异。你要问，“哪台缝纫机是我？”答案很明显，它们当然都不是你——而那是因为，它们都没有感知任何事。在你看来，那些容纳怪圈的大脑跟缝纫机、纸风车和旋转木马一样死气沉沉，没有特质。但有趣的是，那些拥有容纳怪圈的大脑的生命却无法赞同你而承认自己没有特质。他们中有一位坚称：“我就是在这里看着这朵紫色的花的那一个，而不是站在那里喝奶昔的那一个！”另一个人则会坚持认为：“我是喝奶昔的那一个，而不是看那朵花的那一个！”它们中的每一个都坚信自己身在某处，正在看见、听到和体验事情。你拿什么来否定他们的声明呢？

642 我并不想否定他们的声明。他们的声明是完全成立的——只不过它们成立与否，跟容纳怪圈的大脑没有关系。你关注的事情不对。任何一种关于“在这里”“有意识”的声明之所以成立，是因为有某种额外的东西在，那种东西超越怪圈，高于怪圈，让大脑成为灵魂之所。我没办法告诉你那是什么，但是我知道这就是真相，因为我不仅仅是在宇宙中的某处堆积的物理材料。我体验事情，比如花园里那朵紫色的花，和几个街区以外的摩托车的吵闹声。而我的体验才是基本的数据，我所说的

所有其他一切都以此为基础，所以你不能否认我的声明。

641 这跟我刚刚描述的有任何不同吗？一种足够复杂的大脑不但可以感知和划定范畴，还能够用言语表达被它划定范围之物。跟你一样，它也能谈论花朵、花园、摩托车的轰鸣声，它也能谈论自己，说自己在哪里，不在哪里，它还能描述它当下和过去的体验和它的目标、信仰、困惑……你还能要求什么更多的东西呢？它跟你所谓的"体验"难道不一样吗？

642 言辞，全都是言辞！要点在于，体验比言辞包含的东西更多——它还包含感觉。任何称得上是体验的东西，都不得不在看到那朵花灿烂的紫色(注)的同时，还如此感觉到它，而不仅仅是单调地发出"紫色"的声音，就像电话菜单树的自动发音一样。看到一抹鲜艳的紫色的行为发生在言辞、观念或符号的层面之下——它是更加原发的。它是体验者直接感觉到的体验。这就是真正的意识与电话菜单树中机械发音的"人造信号"之间的差别。

641 你会说，不通言辞的动物也能享受这种"原发"的体验吗？奶牛可以跟你一样品味一朵花的深紫色吗？蚊子呢？如果你的答案是"是"，那么这是不是很接近于在暗示奶牛和蚊子跟你有同样多的意识？

642 蚊子的大脑远没有我的复杂，所以它们不可能有跟我一样丰富的体验。

641 等等。你不能二者兼得。就在刚才，你还坚持认为，大脑的复杂性不会造成任何影响——如果一个大脑缺乏那种把有感觉之物与没有感觉之物区分开来的特殊的妙不可言之物（je ne sais quoi），那么它就不是意识之所。可是现在你又说，我们讨论中的大脑的复杂性确实带来了差异。

642 好吧，我想它在某种程度上不得不形成某种影响。一只蚊子不

具备像我一样欣赏紫色花朵的特质。但是也许一头奶牛是具备的，或者至少它是更接近的。但是复杂性本身并不能解释大脑中的感觉和体验。

641 让我们再稍微更深入地思考一下体验和感觉外部世界的这个概念。假设你最喜欢的颜色是紫色，而现在你盯着一张纯紫色的纸片，并让它完全充满你的视野，那么你能体验到在花园里一朵盛开的花瓣上看到紫色时的那种欣然之感吗？

642 恐怕不能。我对于一朵紫色的花的体验之所以那么强烈，有一部分是因为我在每片花瓣上看到的所有微妙的色泽、曼妙的曲线，还有它们围着一个由10多个小黑点构成的发光的中心旋转汇聚起来的样子……

641 不用说，还有这朵花在一根枝条上婷婷而立的方式，而且那根枝条还是灌木丛的一部分，那灌木丛又只不过是五光十色的花园里众多的花丛之一……

642 你是不是在暗示说，我并不是在欣赏紫色本身，而只不过是在欣赏它嵌入一个庞大景象中的方式？这就太离谱了。周边环境也许会加强我的体验，但是我喜爱那种丰满的天鹅绒紫色，纯粹是喜欢那个颜色本身，跟任何其他的什么东西都没有关系。

641 那么你为什么还要用“天鹅绒”这个词来描述它呢？苍蝇或者狗会在紫色的花朵中体验出“天鹅绒”吗？那个词的所指难道不是天鹅绒吗？这难道不是意味着，你的视觉体验唤醒了埋藏在深处的记忆吗？那可能是来自童年的触觉记忆，是你的手指抚过天鹅绒紫色坐垫的感觉。或者你也可能是无意识地记起了自己曾经喝过的一种深色酒，它的标签把它描述成“天鹅绒色”。你怎么能够声称，你对于紫色的体验“跟世界上其他任何东西都没有关系”呢？

642 我想要说的只不过是更复杂的经验都是从基础的、原初的经验

中生长出来的，而即使是这些原初的经验，也跟在简单的物理系统中发生的事——比如绳子在微风中摇荡，或者浮球在抽水马桶中浮沉——存在根本性的和本质上的不同。当微风吹过时，一条摇荡的绳子什么也感觉不到。没有感觉，也没有这里。但是当我看到紫色或者品尝巧克力时，我得到的是一种感官上的体验，而我的心灵生活正建立在数百万种类似的感官体验之上。在这种断裂之间，是一个巨大的谜团。

641 这听起来挺吸引人的，但遗憾的是，我觉得你全都搞反了。那些小小的感官体验之于你心灵生活的宏大模式，就好像小说中的汉字笔画之于小说的情节和人物一样——都只是没有道理可循的不相干的记号而已，并不承载意义。一横一竖都不代表任何意义(注)，而是这些笔画按照复杂的次序联络在一起，就能表达出一部小说或者一篇故事中全部的丰富和人性。

642 这不是谈论故事的正确层面。作家们选用的是字词，而不是笔画，而字词显然是饱含意义的。把很多微小的意义组合在一起，你才能得到一个更大的意义丰富的东西。同样，生命也是由很多微小的感官体验构成的，它们链接在一起，形成一种宏观的感官情绪体验。

641 且慢。孤立的字词是没有深度和力量的。如果把一个字词嵌入到复杂的语境当中，它能够获得巨大的力量，但单独一个字词是没有力量的。认为字词本身具有力量是一种错觉，而认为构成字词的笔画具有力量，是一个更大的错觉。

642 我同意笔画没有力量或意义。但是字词是有的！它们是意义的原子，更大的意义结构都是用它们来搭建出来的。你不可能用没有意义的原子造出意义来！

641 哦？真的吗？我以为你刚刚承认发生在字词和笔画上的正是这么一回事儿。但是没关系——我们换个例子。你觉得音乐有意义吗？

642 音乐是我所知最有意义的东西之一。

641 可是，单独的音符在你看来是有意义的吗？比如，当你听到中央C这个音时，会感觉着迷或者厌恶，美丽或者丑陋吗？

642 我希望不会！不会比我看到单独的字母“C”时感觉到的更多。

641 有没有任何一个孤立的音符，能够凭借它本身迷住你或者让你厌恶？

642 没有。一个孤立的音符不会承载任何音乐意义。说自己被一个单独音符感动的人，都是在装腔作势。

641 可是，当你听到一首乐曲时，不管你喜欢还是讨厌，你总是能被它吸引或者对它排斥。既然乐曲里的任何音符本身都没有让你着迷或者厌恶的性质，那种感觉从何而来呢？

642 这取决于它们在更大的结构中排列的方式。一首旋律之所以吸引人，是因为它具有某种“逻辑”。某些其他的旋律可能让人感到不愉快，这是因为它缺乏逻辑，或者因为它的逻辑太过愚蠢或幼稚了。

641 这听起来显然更像是一种对于模式的反应，而不像是原始的知觉。尽管一首乐曲是由没有情绪意义的声音原子构成，它依然可以获得强大的情绪意义。因此，关键的是组织的模式，而不是构成物的本性。这就把我们带回了你那个关于体验者和非体验者之间差别的谜题，前者比如你和我，而后者则比如摇荡的绳子和塑料浮球。在你看来，这种关键性的差别一定源自某种特别的成分，某种看得见摸得着的东西或者物质，体验者的构成中有这些东西，而非体验者则没有。对不对？

642 差不多是这么一回事儿。

641 那么，让我们给这种使体验者成为体验者的特殊成分起个名

字，叫“感素”（feelium）。遗憾的是，没有人发现过单一的感素原子或者分子，而且我深深地怀疑，就算我们真的发现了这样一种神秘的物质，它只存在于高等生物之中，而不存在于低等生物之中，更不可能存在于单纯的机器里，你还是会纳闷，这样一种无生命、无感觉的单纯*物质*单凭一己之力就引发了感觉，这怎么可能呢？

642 如果感素这种东西真的存在，它可能更像是电流，而不是原子或分子。或者它也可能类似于燃烧或者放射活动——总之是某种看起来活着的东西，某种在本质上疯狂乱舞的东西——而不仅仅是惰性的*原料*。

641 如果你给生物进化前的地球画一幅画，画里会有火山、雷霆、电、火、光和声——甚至太阳，那个伟大的核聚变大球。然而在你的想象里，这类现象的存在，不管通过何种形式的组合或发生何种变异，都永远不可能创造出一个体验者来。可是就在刚在，你在说到那个被我称为“感素”的创造灵魂的神秘物质时，使用到了“舞”这个字，跟“跳舞的符号”这个短语里的“舞”字是一个意思。你是否已经不自觉地改变了自己的思路呢？

642 好吧，我可以想象，把体验者和非体验者区分开来的，是一种闪亮的、火一般的“舞”。在某种意义上，我甚至愿意认为，感素之舞如果真的存在，它也许能够解释体验者和非体验者之间的差别。但是，即使我们理解了感素制造体验的物理机制，某种至关紧要的东西*还是*缺失的。让我们假想，居住在这个世界上的体验者，都是由某种包括感素在内的模式定义的。我们甚至可以假想，就像你说的那样，每个体验者的核心模式都是一个怪圈。就算是这样，那么因为这种让人捉摸不透而又美妙无比的物理模式至少有一部分要在感素中执行，所以在宇宙中每个特定地方的周围，都散布着大量的“亮光”。那个挥之不去的问题依然在：在它们之中，哪一个才是*我*？是什么把它们之中的一个与所有其他区分开来？“我”的来源是什么？

641 你为什么说你跟其他的都不一样呢？每一个都会大喊，它是与众不同的。你们都在异口同声地表达一样的想法。从这种意义上说，你们彼此之间就是无法区分的！

642 我觉得你在取笑我。你很清楚，我跟任何其他人都不一样。我的内在之火在这里，没在任何其他的地方。我想知道，是什么东西把这束特别的火从其他所有的火中分离出来。

641 就跟我之前说过的一样：你是你大脑的一颗卫星。一个特定的大脑跟一个壁炉一样，位于一个特定的地点。不管它到底在哪儿，居住在那里面的怪圈都把那个地点称为“这里”。这里面有什么神秘性呢？

642 你没有回答我的问题。我觉得你甚至都没在听我的问题。

641 哦，当然不是——我在听。我在这里，你在那里！

642 哎哟。你就好好听一下吧。我的问题非常直白。随便哪个人都能理解（也许除了你之外）。我为什么在这个大脑里？我为什么没有落在另外的某个大脑里？比如说，我为什么没有在你的大脑里？

641 因为你的“我”不是一个先在的定义明确的东西，早就羽翼丰满、棱角分明，先天注定要在某个特定的瞬间，跳入某个新造的空空的物理容器当中。你的“我”也不会突然横空出世，毫无预兆，初生即绽放。恰恰相反，你的“我”是从百万个不可预测的事件中缓慢浮现的产物，那些事件都发生在一个特定的身体和那个身体内部的大脑之中。你的“我”是一个不断自我强化的结构，不仅仅是在那个大脑里面逐渐出现，而且还多亏了那个大脑才会存在。它不可能在这个大脑内部存在，因为这个大脑经历了不同的体验，从而造就了一个不同的人。

642 可是你的那些体验，我为什么就不能同样轻易地经历呢？

641 小心！每个"我"都被界定为它的体验者的结果，反过来可不行！把它反过来思考则是一个诱人的陷阱。你在不断地透漏出你没有说出口的一个假设，那就是，任何一个"我"，尽管它是在一个特定的大脑内部成长起来的，却没有深深扎根在那个大脑里——同样的一个"我"，可以同样轻易地在任何一个别的大脑内成长起来，并依附在那个大脑里；在一个给定的"我"和一个给定的"大脑"之间的联系，并不比一只给定的金丝雀和一个给定的笼子之间的联系更深。你可以随便把它们进行更换。

642 你还是没抓住我的要点。我不是在问，我为什么最终落定在这个大脑里，而是在问，我为什么会在那个随机的大脑里开始，而不是其他的大脑。没有任何理由限定，必须得是那个大脑。

641 不，你才没有抓住要点。虽然这可能会让你感到不舒服，但是这里关键的一点在于，没有人会在那个大脑里开始——根本就没有任何人。那个大脑里什么都没有，就跟一条摇晃的绳子或者一个旋涡一样。但是跟那些事物不一样的是，大脑可以感知，可以进化得越来越精密复杂，所以，几周、几个月、几年过后，那里渐渐地会出现某个人。但是那个人的特质不是一下子就完整地出现了；相反，它缓慢地聚合起来，汇成焦点，像天上的云，像窗子玻璃上的凝霜。

642 但是那个人天生注定要成为谁呢？为什么它不能是另外的某个人呢？

641 我马上就要说到了。慢慢占领那个大脑的，是一套复杂的心灵倾向和言语习惯，而且它现在正执拗地重复地问一个问题："为什么我在这里，而不在那里？"你可能注意到了，在这里的这个大脑（也就是我的大脑）并没有一遍又一遍地开口问那个问题。我的大脑和你的大脑是十分不同的。

642 你是不是想告诉我，问"为什么我在这里，而不在那里？"这个

问题，是没有道理的？

641 没错，我是有这个意思，虽然不止于此。所有这一切之所以看上去那么有悖直觉——有时候踩在了不可理喻的边界线上——是因为你的大脑（跟我的一样，跟每个人的一样）已经把一个自我强化的故事讲给它自己上百万次了，那个故事的中心任务被称作“我”，而这个“我”最为关键的一个方面，也是对于“我”而言不可或缺的一个必要方面，就是它频繁无阻地跳到其他的大脑里，至少是有一部分跳了过去。通过亲密关系、同情、友谊、亲缘关系（以及其他方式），你的大脑的“我”持续不断地对其他的大脑发动小小的突袭，在某种程度上通过他们的视角看待世界，因而劝说自己相信，它可以轻松地迁居到它们之中。然后，再自然不过的是，它开始纳闷，为什么自己没有居于它们当中。

642 好吧，它当然会问自己这个问题了。还有比这更自然的问题吗？

641 而有一种答案是，在很小的程度上，你的“我”就居住在别人的大脑里。没错，你的“我”有一部分就居住在我这个愚钝到令人绝望的顽固大脑里，反之亦然。可是尽管这种模糊的外流把界限分明的小城市版本的“你”变成了更大的大都市的“你”，你的“我”还是相当局部化的。你的“我”显然没有均匀一致地散布在地球表面上所有的大脑里——这就好像墨西哥城的都市扩展不可能蔓延到马达加斯加的郊区一样！但是对于你提出的“为什么我在这里，而不在那里？”这个问题，还有另外一种答案，而那个答案会让你十分困扰。那就是，你的“我”没有任何栖身之所。

642 再说一遍？这听起来不像你平时说的话。

641 嗯，这只不过是另一种看待这些事情的方式而已。在更早的时候，我把你的“我”描述成一个自我强化的结构和一个自我强化的故事，但是现在，我要冒着惹恼你的危险，把它称作一个自我强化的神话。

642 一个神话？我肯定不是一个神话，我就站在这里告诉你呢。

641 别着急。先想一想信封盒子里那颗坚硬的玻璃球的错觉。如果我坚持认为那个信封盒子里有一颗真的玻璃球，你会说我从头到脚地掉入一种触觉错觉，不是吗？

642 我确实会这么说，即便感觉到那里面有某种坚硬的东西并不是错觉。

641 同意。所以我的主张是，你的大脑（跟我的和其他人的一样）发明了某种它称之为“我”的东西，这是绝对有必要的，但是那个东西跟信封盒子里那颗“玻璃球”一样真实（或者不如说，是一样的不真实）。在那种意义上，你的大脑欺骗了它自己。那个“我”——你的、我的、每个人的“我”——都是一个无比逼真的错觉，而对这种错觉的迷信，具有奇妙的生存价值。我们的“我”是自我强化的错觉，这些错觉是怪圈带来的不可避免的副产品，而这些怪圈本身又是处理符号的大脑带来的不可避免的副产品。大脑借此引导身体穿过生命的凶险海峡和变幻莫测的水域。

642 你是在告诉我，任何一个“我”都并非真正存在。可我的大脑却同样肯定地告诉我，真的有一个“我”存在。然后你又告诉我，这只是我的大脑在欺骗我。但是，请容我问一句——到底是在欺骗谁？你刚刚告诉我，这个我不存在，所以我的大脑在欺骗谁呢？而且——请再容我多问一句——如果没有一个我作为我的大脑的归属，我怎么能叫它“我的”大脑呢？

641 问题在于，从某种意义上说，一个“我”是某种凭空捏造之物。既然凭空捏造某物是不可能的，这所谓的某物说到底就是一种错觉，虽然是一种非常有力的错觉，比如信封中间的那颗玻璃球。话虽如此，“我”这个错觉远比玻璃球的错觉更加根深蒂固且难以制服，因为在“我”的情况里，没有相当于把盒子倒回来抖一抖的行为，更不能把信

封拨开来看一看，于是发现没有任何坚硬的圆形的东西。我们没有了解我们大脑内在工作机制的入口。所以，我们观察我们的玻璃球“我”的唯一视角只能来自相当于把所有信封一起挤压的动作，而那个视角说，它是真的！

642 如果那是唯一可能的视角，那么我们又能从哪里获得哪怕一丁点儿提示，意识到我们可能在错信一个神话呢？

641 很多人都偷偷地怀疑过，“我”这个概念有点儿神话色彩；引起他们这种怀疑的一个东西恰恰是你在我们的谈话中一直心心念念的东西——也就是，在物理学的铁律和被称作“我”的朦胧模糊的东西的存在之间，似乎有某种不和谐之物。在只有无生命的东西来回移动的世界里，怎么可能存在体验者呢？感知、感觉和体验似乎是某种额外的东西，超越物理，在物理之上。

642 当然，除非有感素这种东西存在，可是这丝毫都不明朗。不管怎么说，我都同意，“我”这个概念跟物理学发生的冲突暗示它是一个非常难以捉摸的东西，而且非常需要一个解释。

641 这里还暗示有某些东西需要修正，跟我们感知到什么引发了什么有关系。在我们的日常生活中，我们想当然地认为一个“我”能够引发事情，并支配这些事情。如果我决定开车去食品杂货店，我那辆一吨重的汽车就会把我载到那里，再把我载回来。这在物理世界里看起来就相当诡异了，因为在那个世界里，每件事都只是粒子交互的结果。这个粒子主导的故事如何能给一个朦胧而缥缈的“我”留下空间，让它引发一辆沉重的汽车移动到某处呢？这也为“我”这个概念的真实性蒙上了一层疑团。

642 可能是吧——但即便真是这样，它也很轻很轻。

641 再轻也不要紧。那层极轻的疑团悍然违犯了我们从最早的童年

时代就已经视为理所当然之义，即“我”是的确存在的——在绝大多数人那里，后一种信念都轻而易举地胜出了，兵不血刃。在大多数人的思维里，这场战役从来就没打响过。人们找到了各种各样的安全舱口，包括意识是一种新型量子现象的概念(注)，以及意识均一地存在于所有物质之中的观念，等等。我为了结束这场战争而提出的休战提议，是把“我”视为由幻觉感知到的一种幻觉。这听起来相当奇怪，甚至更奇怪一点地说：“我”是被幻觉幻想出来的一种幻觉。

642 这听起来已经不仅仅奇怪了。简直是疯狂。

641 可能吧，但是跟很多现代科学的奇怪结论一样，它可以听上去很疯狂，但还是正确的。曾几何时，有人认为地球在运动而太阳是静止的，而这听起来就很疯狂，因为事实显然是相反的。今天，根据不同的情况，我们能用两种视角来看待这件事了。当我们处于日常的思维框架中时，我们说：“太阳落山了。”当我们在用科学的思维框架思考时，我们会想起来，其实只是地球在转动。我们是灵活的生物，有能力根据不同的情况转换我们的视角。

642 所以，根据你的观点，我们是不是也应该有能力转换视角，看到一个“我”的存在呢？

641 当然了。我说一个“我”是由幻觉感知到的一种幻觉，有点像是日心说的视角——它可以提供新的洞见，但是非常有悖直觉，而且也很难跟他人顺利交流，因为大家对于他们的“我”的信仰，都抱有一种不可战胜的狂热。我们解释我们自己的行为和他人的行为，是通过设立我们自己的“我”和它在他人中的类比物来进行的。这种天真的视角让我们能够用人们可以理解的方式谈论人的世界。

642 天真？！我注意到，你从未停止过说“我”！你可能在过去的五分钟里已经说了100次“我”了。

641 千真万确。你说得一点儿也没错。这个“我”对我们所有人而言的都是一个必要的、不可或缺的概念，即便它是一个错觉，就好像我们认为太阳绕着地球转是因为它升起来，穿过天空又落下去一样。只有当我们观察“我”的天真视角与物理世界发生冲撞时，它才遭遇了各种各样的困难。正是在那一点上，我们当中倾向科学的那些人才意识到事情可能还有另外一副面孔。但是，相信那个关于“我”的简单说法对于大多数人来说，都比为“我”找到一个科学的解释重要百倍，所以结果就是，二者之间根本就没有竞争可言。“我”的神话轻松胜出，从未听见过辩论的声音——甚至在绝大多数倾向科学的人的思维里，也未曾发生过争辩！

642 那怎么可能呢？

641 我推测有两个原因。第一个原因是，“我”的神话比“太阳绕着地球转”的神话更加无限地接近我们信仰系统的中心；而第二个原因是，任何一种对于它的科学性另解，都比向日心说的转换更加微妙、更令人迷惘。所以，“我”的神话比“太阳绕着地球转”的神话更难从我们的思维中分离出来。让一个成年人放弃“我”的概念，差不多跟让一个小孩子不再相信圣诞老人一样困难。实际上，放弃圣诞老人的故事与放弃“我”的执念比起来，不值一提。完全停止相信“我”其实是不可能的，因为它对于生存是不可或缺的。不管我们喜欢也好，讨厌也罢，人类就是要永远身陷于这个神话之中。

642 为什么你非要坚持说，“我”跟那颗受到诅咒的玻璃球一样，只是一种神话、幻觉或者错觉呢？你老是炫耀那个老掉牙的玻璃球隐喻，我真是受够了。我想知道的是，被幻想出来的错觉到底是什么？

641 好吧，我们先让那个玻璃球隐喻歇一会儿。基本的观念是，大脑中的符号之舞本身就在被符号感知，而感知这一步又延长了那舞蹈，如此旋转不息。简单地说，那就是意识。但是如果你还记得符号只不过是非符号化的神经活动制造的宏观现象而已，所以你可以转换视角，完

全摆脱符号的语言，这样一来，“我”就分崩离析了。它就那么“噗”地一下子，消失不见了，所以也没有给向下的因果性留下任何空间。

642 再说得具体一点儿，这是什么意思？

641 意思是，在这幅新的图景里，没有欲望、信念、人物特征、幽默感、观念、记忆或者任何一种心灵性的东西；留下的只有零碎的物理事件（实质上就是粒子冲撞）。人们在闹壳里也可以做同样的事，可以转换视角，要么在小磁交弹球的层面上看待事物，要么在小磁交弹的层面上观察。在前一种层面上，小磁交弹是完全看不见的，而在后一种层面上，小磁交弹球则是隐而不显。这两种竞争的视角确实是对立的两极，跟地心说和日心说一样。

642 这些我都懂，但是你为什么要不断地暗示说，这两个观点中有一个是错觉，而另一个是真相呢？你总是把优先的地位让给粒子的视角，那个更低层级的微观视角。你的偏见为什么这么深呢？你为什么不能简单地看到两个同样良好的竞争性观点呢？我们可以根据自己认为合适的情况，在二者之间来回摇摆。这有点儿像物理学家在处理气体时，在热动力学和统计力学之间的来回摇摆。

641 因为，无比遗憾的是，非粒子的观点包括了好几种魔法式的思想。它把世界划分成两种根本不同的实体（体验者和非体验者），包括了两种根本不同的因果性（向下的和向上的），还包括了那些凭空闪现而又在某个时刻突然熄灭的非物质的灵魂，等等。

642 你可真是前言不搭后语！你喜欢借用641的质数性来解释多米诺骨牌的翻倒！你偏爱这种解释！你曾经坚持说，这才是那张多米诺骨牌没有翻倒的真正原因，而另一种解释方式是短视的，无用到无可救药。

641 正中要害！我承认，我的立场带有一丝反讽的色彩。有时候，

严格的科学观点就是无用到无可救药的，即使它是正确的。这就是一个两难的困境。正如我之前所说，人的境遇，从本性上来说，就是对于神话的笃信。我们将永恒地身陷于那个境况之中，生命也因此得以保持相当的乐趣。

642　道家和禅宗在很久以前就意识到了这桩事情自相矛盾的状态[注]，而且都特别努力地想要拆除、解构或者干脆摆脱那个“我”。

641 听起来好像是个崇高的目标，但是它注定会失败。正如我们需要我们的眼睛来看见一样，我们也需要我们的“我”去存在！我们人类的命运就是有能力感知抽象物，而且有动力去感知。我们这些生命一生都在把世界归类到不断增长的模式等级秩序当中，而这些模式全部是由我们大脑中的符号来表征的。我们通过把先前的符号在新的结构中组织起来，不断创造出新的符号，几乎无穷无尽。另外，站在宏观层面的我们看不到下面的物理因果性发生的层面，所以作为一种补偿，我们发现了各种各样非凡有效的速记式方法来描述发生之事，因为这个世界虽然相当疯狂而凌乱，可仍然充满了我们在大多数时候可以信赖的法则。

642 你说的哪种法则？

641 哦，举个例子吧。当你在操场上荡起秋千时，它会以一种能够预见的方式悠荡，尽管它的链条和座位的运动细节远远超过了我们的预知能力。我们丝毫不关心那个层面上的细节。我们感觉我们已经极好地掌握了秋千的运动。同理，购物手推车在你的推动下基本上会按照你想让它走的路线前进，尽管它们摇摇晃晃的轮子会画出一条不可预知的随机的轨迹。在人行道上散步的人可能会做出某些不可预料的动作，可是你完全不用担心他们会变成一个巨人，把你一口吞下去。这些法则为我们深谙于心，并视为理所当然，而它们与粒子冲撞的层面相距甚远，令人吃惊。一切之中最有效、最令人难以抗拒的速写法，就是把抽象的欲望和信念归因于某种“特权”实体（那些有灵魂的——动物和人），还有

那将所有东西整合为单一的、按理不可再分的统一体，且代表这样一种实体的“核心实质”。

642 你的意思是，那个实体的“灵魂”？

641 差不多。或者，如果你不愿意用那个词的话，我们还可以说，那是你假想中那个东西内部感觉的方式——它的内在视角。然后，最要紧的是，既然每个感知者都一直浸淫在它自己的行为和数不清的后果当中，它难免要为它自己的灵魂，亦即它自己的核心实质，编造出一套格外精密的说辞。这套说辞跟它给其所见的拥有思维的其他实体编造的说辞没什么两样——只是在细节上远远高出而已。不仅如此，有关一个“我”的故事是一套关于核心实质的说辞[注]，而那种核心实质从不会消失在视野之外（与“你的”、“她的”和“他的”不同，它们可能在一两个场景中露面，然后便退回幕后）。

642 所以，那让这个系统注定陷落在这种错觉之中的事实，便是它能够观看它自己。

641 不仅仅是它能够观看自己，而且它真的在观看自己，而且一直如此。另外还有一个关键性的事实，是它除了从根本上简化每一件事之外，别无选择。我们的范畴是对于世界上各种模式的巨大简化，但是那些精选的范畴无比有效地让我们得以思考和预料自己周遭世界的行为。

642 我们为什么不能摆脱我们的幻觉呢？我们为什么不能抵达禅宗人士追求的纯净而没有“我”的状态呢？

641 我们可以尽管尝试，在短时间里这也是一项有趣的练习，但是我们要想在世界上存活下去，就不可以关掉我们的感知机器。我们没办法让我们自己不去感知如树、花、狗和他人这样的东西。我们可以玩这场游戏，可以告诉自己我们成功了，可以声称我们已经对它们“去感

知”了，可那都只是显见的自欺欺人而已。事实是，我们都是宏观层面上的生物，所以我们的感知与我们的范畴，相对于宇宙的真正因果性所处的构造而言，都是极其粗线条的。我们被困在了根本性简化的层面上，不论这是好事还是坏事。

642 那岂不是一场悲剧？你让它听起来像是一种悲伤的命运。

641 完全不是——这是我们的光荣！只有那些笃信禅宗和道家的人才会十分严肃地认为这是一个需要我们与之战斗的境况，就像针尖对麦芒。他们憎恶言辞，也憎恶把世界打破成独立的碎块，并为它们命名。所以他们会给你提供配方——比如他们玄妙的心印（koans）——以挑战这种普世固有的使用言辞的驱动力。我自己在理解世界的谜题时，没有任何对抗这种言辞使用的欲望——真实情况恰恰相反！但是我承认，言辞使用具有一个十分重大的缺点。

642 那是什么？

641 那就是，我们不得不忍受悖论，还要以最亲密的方式与它共存。而“我”这个词是所有一切的集中体现。

642 我看不出“我”这个词有任何哪怕是一点儿的悖论性。事实上，我看不出在“我”这个平凡而直白的概念和哥德尔式怪圈这样一个深奥难解、几乎不可能把握的概念之间有任何类比性。

641 好吧，这么想一想。一方面，“我”这一表达指的是一套高度抽象的概念：一段人生故事、一种品位、一簇希望与恐惧、某些天赋与缺陷、特定程度的机智、不同程度的心不在焉等等。然而在另一方面，“我”这一表达还指一个由亿万个细胞组成的物理实体，每个细胞都在做自己的事情，对于所谓的“整体”没有丝毫的关切，它们只不过是这个“整体”最微小的组成部分。换句话说，“我”同时在指一个高度可感而有形的生物基底和一个高度不可感且抽象的心理模式。当你说“我饿

了”的时候，你指的是这两个层面中的哪一个呢？而当你宣布“我很幸福”的时候，你又是在指哪一个呢？当你坦白“我不记得我们之前的电话号码了”的时候呢？当你欢叫“我爱滑雪”的时候呢？当你打着哈欠说“我困了”的时候呢？

642 没错，既然你现在提到了这一点，我的确同意，“我”所代表的东西是有点儿难以锚定。有时候，它指的是具体的物理实体，而有时候是抽象的心灵实体。可是当你归结起来看时，“我”总是既具体又抽象的。

641 它只是用两种在现象上不同的方式来描述的同一种东西而已，而那正好跟哥德尔的句子一样。这也就是为什么我们可以理直气壮地说，它能够同时谈论数字和它自己。同理，“我”也是同时在说一群数不清的分离的物理实体和一种抽象的模式——正是这种模式说出了这个词！

642 看起来，这个小小的代词似乎就是让我们人类的存在显得神秘而神奇的枢纽所在了。它和周围其他的东西太不一样了。“我”这个代词中包含的那个于内在指向自我的环路[注]——哲学家们会把它叫作“索引性”[注]——与宇宙中所有的其他结构在本质上都有所不同。

641 我不太同意。或者说，我完全不敢苟同。“我”这个代词并没有包含比哥德尔建构核心处的自我指涉更强、更深或者更神秘的一种自我指涉。事实恰恰相反。只不过，哥德尔把*我*真正的意思清楚地说了出来。他揭示出，在所谓的“索引性”场景幕后，其实只有编码和依靠着稳定、可靠的类比系统而产生的对应关系。被我们称为“我”的那个东西，就来自那种指涉的稳定性，仅此而已。“我”跟任何其他有所指的单词相比，都没有任何更加神秘之处。如果说有什么神秘东西的话，*语言*才是那个跟宇宙中其他结构都不一样的东西。

642 所以在你看来，我一点儿也不神奇？存在一点儿也不神秘？

641 我没有这么说。我感觉存在非常神秘，因为，跟每个人一样，我是有限的，我没有能力看到我基底足够深的地方，让我的我噗的一下消失不见。如果我可以做到的话，我猜生命将会变得相当无趣。

642 我应该想到的！

641 当我们通过科学实验，真正向下看到我们清晰的基底时，我们会发现跟我具有同等哥德尔性的小小的奇迹。

642 啊，没错，当然了——小小的微型哥德尔精灵！但是……比如呢？

641 我是指DNA双螺旋的自我复制。它背后的机制都涉及与哥德尔式自我指涉中蕴含的相同的抽象观念。这就是约翰·冯·诺依曼在20世纪50年代早期设计出一台自我复制的机器时于不自觉中揭示的道理[注]，而它的机器也具有跟哥德尔的自我指涉同样的抽象结构。

642 你是不是在说，微型哥德尔精灵是自我复制的机器？

641 正是！这是一个微妙而美丽的类比。哥德尔配数*k*的类比是一个特别的蓝图。“父母机”检查这份蓝图并严格遵从它的指令——也就是说，它会建造蓝图描画之物。为了做到这一点，机器不得不知道每种图标代表什么东西——一种哥德尔式的编码，或者说是映射。新建造的东西是一台机器，但还缺少一个关键的部分。为了填补这块缺陷，父母机就要复制这张蓝图，然后把复本（也就是关键的缺失部分）粘贴到新机器上，然后，请看！——新合成的东西成为一个“孩子”机，跟它的父母一模一样。

642 这让我想起了莫顿盐的标志。“孩子机”是不是跟空手站在那里

的“打伞女孩”缺少同样的关键部分？那张蓝图就是那只小小的蓝色盒子？

641 正确！把那个小蓝盒子给她，她就放飞了！看到了吧，无限！而且令人惊奇的是，就在几年之后，分子生物学家就发现，约翰·冯·诺依曼的哥德尔式机制跟事物自我繁殖的自然规律一模一样。当然，DNA就是那张蓝图。一切都系于那个稳定的映射关系的存在（此处即“基因编码”）和来自它们的意义。看看这带来了什么吧。到目前为止，它造就所有的生命，而且还将带来此后所有的生命！看到了吧，无限！

642 所以你主张的是，在捉摸不定的“我”这个单词魔法般的索引性中反映出来的对于自己身为一个独特活物的意识，其实并不是一种意义深远的现象，而只不过是由映射带来的寻常结果而已？

641 我不认为我说过这样的话！意识到自己活着，意识到自己是无限长的链条上独一无二的一环，这当然*是*意义深远的。只不过，它并没有超越物理法则。相反，它是一次意义深远的对于物理法则的运用——很难用寻常来形容它！另一方面，还有一种过于常见的欲望，试图把“我”这代词神秘化，想让它看起来仿佛隐藏着一个比其他词汇更深的秘密，这确实是把水给搅浑了。所有这些奇怪现象的唯一根基就是感知，是它把符号和意义带入了物理系统。感知，是从威廉·詹姆斯的“巨大的、盛开的、嗡鸣的混沌”向一个抽象的符号层面完成了一次神奇的跳跃。然后，当感知终于不可避免地扭转回头而专注自身时，你就会得到丰富的魔法一般的结果。注意，是魔法*一般*的，而不是*真正*的魔法。你得到了一个跨越层级的反馈环，而它显然的坚实性击败这个了世界上所有其他事物的真实性。这个“我”，在思维之中并不真实可又顽固无比的玻璃球，这种“副”现象，一跃而成为主宰，把自己钦定为“现实第一号”，而且牢牢守住自己的宝座，不管人们怎么说，始终都不肯让位。

642 所以“我”太过玻璃球化了[注]——甚至到了油盐不进的地步？

641 什么？！我还以为你觉得我关于“我”的观念都是天方夜谭。

642 没错，我是这么觉得过，但是我想我现在有点儿跟上你的思路了。可能我绕过来了一点儿。你把一个“我”当作怪圈的观点近乎是自相矛盾的，可是又不是。它很像艾舍尔的《画手》——当你被它奇妙的真实性带到那幅画的内部时，它看起来是自相矛盾的，可是当你退后一步，从外部来看它的时候，这种自相矛盾性就消失了。可那就变成另一幅画了！真是妙不可言。这些太过分了，或者说，就是太像贝里了……绝不可能在罗素的字典里出现。

641 啊，真是句句入耳，字字落心！你能在我的观念中发现一些优点，我真是开心。如你所知，它们只不过是隐喻而已，但是它们帮助我弄明白了关于活着的巨大谜团，还有你在不断强调的，身在这里的巨大谜团。感谢你给了我这次千载难逢的好机会，跟我交流关于这种微妙事物的观点。

642 说真的，千万别客气。我将带着欣悦、快意、殷切、活力、精力、生气、急智和迫不及待的心情[注]，期待我们下一次见面。到时再见了。回见！

·演员退场·

---

1. See [Monod], [Cordeschi], and [Dupuy 2000] for clear discussions of the emergence of goal-orientedness (*i.e.*, teleology) from feedback.
2. See Chapter 22 of [Hofstadter 1985] for a discussion of the abstract essence of hurricanes.
3. See [Hardy and Wright] and [Niven and Zuckerman] for this classic theorem, the simplest case of Waring’s theorem.
4. See [Chalmers] for a spirited defense of the notion of qualia, and see [Dennett 1991], [Dennett 1998], [Dennett 2005], and [Hofstadter and Dennett], which do their best to throw a

wet blanket on the idea.

5. See the dialogue “Prelude… Ant Fugue” (found in both [Hofstadter 1979] and [Hofstadter and Dennett]) for a discussion of how meanings at a high level can emerge from meaningless symbols at a low level.

6. See [Penrose], which views consciousness as an intrinsically quantum-mechanical phenomenon, and [Rucker], which views consciousness as uniformly pervading everything in the universe.

7. Far and away the best book I have read on these spiritual approaches to life is [Smullyan 1977], but [Smullyan 1978] and [Smullyan 1983] also contain excellent pieces on the topic. These ideas are also discussed in Chapter 9 of [Hofstadter 1979], but from a skeptical point of view.

8. See [Dennett 1992] and [Kent].

9. See [Brinck] and [Kent].

10. 索引性（indexicality），是语言哲学中谈论语言表达时使用的一个区分性术语，形容在指称某物时无法独立锁定所指，需要与语境和环境结合起来确定所指的性质。例如，“你”“昨天”，都是索引式的表达，而“美国”“月球”则几乎不具有索引性。——译者注

11. See [von Neumann] for a very difficult and [Poundstone] for a very lucid discussion of self-replicating automata. See Chapters 2 and 3 of [Hofstadter 1985] for a simpler discussion of the same ideas. Chapter 16 of [Hofstadter 1979] carefully spells out the mapping between Gödel’s selfreferential construction and the self-replicating mechanisms at the core of life.

12. 借用了由强尼·莫瑟和理查德·怀丁共同创作并由法兰克·辛纳屈完成后无来者的演唱的一首情歌中的几句话。

13. 我父亲的朋友鲍勃·赫尔曼（一位顶尖的物理学家，因为在宇宙背景辐射被观察到的15年前便与人共同预测了它的存在而知名）很喜欢背诵这个谜语，还要模仿出一种很重的意第绪口音：“流浪汉在树林里捅了一个马蜂窝。当马蜂带着欣悦、快意、殷切、活力、精力、生气、机智和迫不及待的心情叮咬他时，‘噢’，他在心里想，‘如果我右边扁桃体左侧上的包的数量是在致命点和喉结之间的包的八分之七的6.75倍，那么如果一匹马有6条腿，一个在向下移动的楼梯上向上滚圈的男孩要花多长时间才能数完木板上的石子？’”。所以我觉得我理应向鲍勃致以迟到的敬意。

第21章 与笛卡儿自我发生一场短暂的小摩擦

好故事弹出有力的和弦

在前面的对话里，怪圈642号坚持不懈追问的问题是：“是什么让我居于这个特定的大脑之中，而不是任何其他的大脑？”然而，即便怪圈641号努力用好几种不同的方式为这个谜题提供答案，怪圈642号总是觉得怪圈641号并没有真正理解这个问题，而且也没有理解它对于人类的存在具有多么深刻的核心意义。有没有可能，这里面出现了一个交流中的根本性断裂，有些人就是永远也没办法理解这个问题，因为它太微妙、太捉摸不定了？

好吧，如果我们对于科幻剧情并不排斥，那么同样的问题就可以表达得生动而直白，这样一来，人们有望更好地理解这个难题，而且不再陷入深深的困扰。在牛津大学哲学家德里克·帕菲特（Derek Parfit）有开拓之功的著作《理与人》（*Reasons and Persons*）里，就演示过一种如此做的方式。帕菲特设定谜题的方式如下所示：

> 我进入电子传输器。以前我去过火星，然而却只是借助那种老的方式——花费数周时间乘坐太空飞船旅行的方式——前往的。而这部机器将以光速把我送达火星。我需要做的只是按一下那个绿色的按钮。像其他人一样，我感到紧张。能行吗？我提醒自己人们所告知我的预期的情形：当我按下按钮时，我将失去知觉，然后似乎在片刻之后会苏醒过来。而事实上我将失去知觉大约一个小时。在记录我的所有细胞的精确状态的同时，地球上的这个扫描仪将毁坏

我的大脑和躯体。然后它将通过无线电信号传输该信息。

信息以光速运行，将用三分钟的时间到达火星上的复制器。然后这部机器将用新的物质创造出与我的大脑和躯体精确相似的一个大脑和躯体。我将在该躯体中苏醒过来。

尽管我认为这就是将会发生的事情，但还是有些犹豫不决。好在接着我记起了今晨早餐时，我对此流露出紧张情绪的时候，妻子粲然一笑的情景。当时她提醒我说，尽管她经常被电子输送，<u>她</u>还不是好好的。我于是按下按钮。正如所预期的，我失去了知觉而且好像立刻重获知觉，但却身处一个不同的舱室中。审视着自己的新身体，我发现一点也没变。甚至早晨刮胡子时在上嘴唇上留下的那个小伤口也仍然在那里。

*

几年过去了，期间我时常被电子传输。现在我又一次步入那个舱室中，准备再一次到火星上去。但是这一次当我按下绿色按钮时，我没有失去知觉。先是出现了嗡嗡的声音，然后是一片寂静。我离开舱室，对服务员说："机器没有正常工作。我的操作有什么不对吗？"

"机器工作正常，"他一边回答我，一边递给我一张打印出的卡片。卡片上写着："新扫描仪记录下了你的图像，没有毁坏你的大脑和躯体。我们希望你会欢迎这个技术进步所带给你的机会。"

服务员说我是首批使用新扫描仪的人之一。并且补充说，如果我逗留1小时的话，我就能用内部通话设备看到并与火星上的我自己对话。

"先等一下，"我回答说，"如果我在这里，我就不能<u>也</u>在火星上。"

有人轻轻地咳嗽一声。这是一个身穿白大褂的人，他要与我私

下谈谈。我们来到他的办公室。他让我坐下，停顿了片刻后说："恐怕我们的新扫描仪遇到了一些问题。它就像你与火星上的你自己交谈时所看到的那样精确地记录了你的图像。但是看来它有损它所扫描的对象的心脏系统。就迄今的结果来判断，尽管你将在火星上相当健康地生活，但是在这里，在地球上，你不日将难逃心力衰竭的厄运。"

后来服务人员把我叫到内部通话设备旁边。在内部通话设备荧屏上，我看到了自己，就像每天早晨在镜子中看到自己一样。但是有两个不同之处。我在荧屏上不是左右反相的。而且，就在我站在这里默默无言时，我能够看到并且听到我自己在火星上的工作室内开始对我讲话。

由于我的复制品知道我将要死去，他就用我最近用来安慰一个垂危朋友的那些同样的方法来努力安慰我。得悉这些想法在接收端多么没有安慰后果真是件伤心的事情。然后我的复制品向我再三保证他将从我生命停止的地方延续我的生命。他爱我的妻子，并将一道照顾我的子女。而且他将完成我正在写的这本著作。除了拥有我所有的草稿之外，他还具有我所有的意向。我必须承认，他能够像我一样好地来完成我的著作。所有这些事实对我起到一点安慰作用。在我知道自己将有一个复制品的情况下慢慢死去，并不太像简单地死去那么糟。即便如此，我还是将很快失去知觉，永远地失去知觉。

我们是多么没有主见的人啊！

帕菲特这则分成两部分的故事，围绕着特别的担忧而展开，那正是困扰怪圈642号的那些担忧。在第一部分，帕菲特在地球上被打碎成原子，而载着他超级细化构造图的信号在抵达火星后又为他重造出一个新的身体，我们跟着他一起担心，在这之后，他是否还能真的存在；我们害怕这个新造出来的，只不过是某个长得跟帕菲特一模一样且思想也跟

帕菲特一模一样的人，但是并不是帕菲特本人。然而，没过多久，我们就得知自己的担忧是没有根据的，因而大松了一口气：帕菲特本人幸存下来了，毫发无损。太棒了！我们是怎么知道这一点的呢？因为他是这么告诉我们的！但是告诉我们这个好消息的，是哪个“他”呢？是这位哲学家和此书作者德里克·帕菲特，还是那个大无畏的空间旅行者德里克·帕菲特呢？

是空间旅行者帕菲特。真实的情况是，哲学家帕菲特只是在编造一个好故事，尽力想让它听起来像教科书上写的一样真实，可是我们很快就发现，在他自己讲的故事里，有好几个部分都是连他自己也不相信的。他幻想的第二个场景是从对第一个场景的颠覆开始的。我们发现，新的扫描仪同旧的扫描仪不一样，它并不毁灭“原来的”，于是我们马上就心照不宣地萌生了这样一种念头：大无畏的太空旅行者帕菲特，其实没有到任何地方旅行。我们不会质疑从地球上的小隔间里走出来的他，因为他还在这里。

噢，可是我们是多么不走心的没有主见的人啊！我们对场景一的“心灵传输等于旅行”主题照单全收，可是在场景二中，我们则好像不假思索地就选了一条阻力最小的途径，差不多如下所述：“如果有两个不同的东西，都跟帕菲特长得一样，想的一样，说的一样，又如果其中之一身在我们上次看到帕菲特的地方，而另一个则身处很远的地方，那么，老天爷做证，离我们近的显然是真的那一个，而离我们远的只不过是一个复制品——克隆人、伪造品、冒充者、假货。”

上面这些已经足够让我们思考好一阵子了。如果场景二中火星上的复制品是一个假货，那么为什么场景一中的那个复制品不是呢？为什么我们在阅读场景一的时候就没看出来呢？我们天真地被他的妻子在早餐时的微笑所收买，然后，当他从火星的小隔间里走出来时，他脸上那道标志性的疤痕打消了我们所有的疑虑。我们轻信了他的话，相信从那个小隔间里走出来的确实是他。但是我们还能期待什么别的吗？那个新生

的身体难道会走出隔间之后大喊："啊，好恐怖，我不是我了！我是另外一个人，只是看上去像我而已，而且拥有我童年以来的全部记忆，甚至我在不久之前刚刚跟我的妻子共进早餐时的记忆！我只是一个赝品，可又是多么逼真的一个赝品啊！"

那个刚刚造出来的火星人当然不会说这种自相矛盾的话，因为他不可能知道他是假的。他会全心全意地相信他就是那个原来的德里克·帕菲特，不久之前刚刚在地球上的扫描仪里被分解了。说到底，这就是他的大脑会告诉他的内容，因为这个大脑跟德里克·帕菲特的大脑是完全一致的！这一切都表明，我们在处理有关个人同一性的声明时，哪怕面对的是从第一人称口中直接说出来的话，都要极其小心。

那么，鉴于我们这个新的并非毫无意义的态度，我们应该怎么思考场景二呢？我们得知，本来要成为太空旅行者的帕菲特却从地球上的隔间里走了出来，而且心脏受到了损坏。但是我们怎么能知道，这一个就是帕菲特呢？为什么给我们讲这个故事的帕菲特没有采用那个也叫自己"德里克·帕菲特"的新火星人的有利视角呢？设想这个故事是以下面这种方式讲给我们的："我刚从火星上的隔间里走出来的时候，就听说了那个可怕的消息，另一个帕菲特——那个留在地球上的可怜家伙——在把我传送到这儿的过程中遭受了严重的心脏损害。我听到这个消息，心都碎了。很快，他和我就要通电话了，而我发现自己处在了一个十分古怪的位置上，要用我最近安抚一位即将去世的朋友的方式来安慰他……"

如果这份重述足够圆滑流畅，那么我们可能无法不去想，也许这个身体，这个存在于火星上的躯体才是真的德里克·帕菲特。确实，德里克·帕菲特作为一名成熟的哲学家和技艺精湛的讲故事的人，也许真的能让我们想象出地球上那个心脏受损的身体只是一个假扮者，装作是通过出身与神圣的法令同"德里克·帕菲特"这个名字联系在一起的那个独一无二的灵魂。

一个横跨大西洋两岸心灵传输的思想实验

看起来，以何种方式讲述一个科幻情节，左右了我们在直觉上对其可信度的判定。对于这一点，我的前同事和老朋友丹尼尔·丹尼特已经在他关于哲学家们巧设的思想实验的讨论中表达过很多次了。丹把这些周密设计、精巧谋划的寓言称为直觉泵[注]，而且他对于自己说的是什么再清楚不过了，因为他本人就曾在心灵哲学的领域里编写过几个最具启迪性的直觉泵。

而且，我不得不说的一件事是，当我在把帕菲特的故事从他那本1984年出版的书中抄录到这一章里的时候，我心里有个声音在小声地对我嘀咕说："你说说，这有没有让你想起丹给《心我论》（*The Mind's I*）写的前言？这本书在1981年面世的时候，他那个天才的心灵传输故事吸引到了那么多的读者。"所以，等我把帕菲特的故事录入完毕后，我从我的书架里抽出了一本《心我论》，重读了它的前几页。让我惊讶的是那完全是同一个故事，只不过星球和性别颠倒了过来，而且讲述的方式更具美国风格而已。他的版本具有同样的二分结构，第一部分塑造了一个"心灵克隆标准四代"（Teleclone Mark IV），它会毁掉原来的身体，而第二部分则描述了一个保留原来身体的最新升级版本（"标准五代"）。

我能说什么呢？这两个故事来自大西洋两岸，不论其中一个是另一个的"克隆体"，还是说它们两个有各自不同的出处（这似乎不太可能，因为《心我论》出现在了帕菲特的参考文献目录里），我两个都爱。不管怎么说，既然我已经把这点小插曲交代完了，我将继续评论帕菲特那个充满挑战性的故事（当然，得益于类比的指涉力量，我同时也是在评论丹的故事）。

笛卡儿自我的不明下落

帕菲特的故事提出的关键问题是："在场景二中的心灵传输完成后，空间旅行者德里克·帕菲特真正在哪儿？"换句话说，哪一个声称自己是帕菲特的人才真的就是帕菲特？在场景一中，讲故事的帕菲特留下了一个几乎显而易见的答案，可是等到了场景二里，他又在几乎同样显而易见的程度上颠覆了那个答案。此时，你大概可以亲耳听见怪圈642号与太空旅行者形成的强烈共鸣，以及他的大声疾呼："我到底是这两个中的哪一个？"

在我看来，如果一个人不能为这个看起来极其自然而又亟待解决的问题提出（或者支持）某种答案，那么就不能声称自己在意识的谜题上有任何重要的观点。我想，到目前为止，你应该已经知道我对于这个问题的答案了，但是也有可能你还不知道。不论如何，我先让你在这个问题上多思考一会儿，与此同时，我继续给你讲讲，帕菲特大致上是如何看待这个问题的。

这个议题是帕菲特那本书的核心，他为了解释自己的立场，用了100页的篇幅。他把他反对的一个关键概念起名叫"笛卡儿式的纯粹自我"，或者也可以简化成"笛卡儿自我"（Cartesian Ego）。用我的话说，一个笛卡儿自我构成了一个纯粹的灵魂量子（另一种说法是"个人同一性"），而且它是百分之百不可切分也不可稀释的。简言之，它就是让你成为你而我成为我之物。我的笛卡儿自我是我的，而不是任何其他人的，从我出生时开始存在，到我死亡时消失，仅此而已。它是属于我自己的，是一个完全私有的、不曾分享也无法分享的第一人称世界。它是我所有经验的主体，是我绝对独一无二的内在之光。你知道我的意思！

顺便插一句，我不得不承认，每次看到"笛卡儿自我"这个短语时，虽然我的眼睛只感知到一个"g"，有一部分我总是幻想出另一个"g"，于是一个鸡蛋（egg）的形象便在我的大脑中闪现出来——如果能允许我这么说的话，就是一个"笛卡儿蛋我"（Cartesian Eggo）：一个形状完美

的鸡蛋，裹着一尘不染的白色蛋壳以保护核心的蛋黄，它具有完美的球形，无限宝贵。在我幻想出来的奇怪而扭曲的画面中，那个蛋黄就是人类同一性的秘密——哎哟，这样一来，帕菲特写作这本书的核心使命就成了无情地打碎整个鸡蛋，连同那神圣的蛋黄一起粉碎！

帕菲特尽全力回答的问题有两个。第一个是：当帕菲特在场景一中被心灵传输到火星上时，他的笛卡儿自我也跟着他一起被心灵传输过去了，还是跟着他的身体一起被粉碎了呢？第二个问题看起来更加急切，也更令人不解，那就是：当帕菲特在场景二中被心灵传输到火星上时，他的笛卡儿自我到哪里去了？它有可能抛弃了地球上的他而来到了火星吗？如果是那样的话，留在地球上的那个人是谁？或者反过来，帕菲特的笛卡儿自我是不是原封不动地留在了地球上？如果是这样的话，从火星的隔间里走出来的那个人——如果这个人存在的话——又是谁呢？（注意，我们把“谁”这个字或者“是谁”这个短语跟一种特定的、具有唯一可辨识性的笛卡儿自我合并在了一起。）提出这些问题（并且相信这些问题有客观的正确答案）的冲动几乎是无法抑制的。然而，引发这种冲动的普遍直觉正是帕菲特想要在他的书中摧垮之物。

说得更具体一点，帕菲特坚定抵制“个人同一性”的概念是有意义的这种观念。当然，在我们栖身的日常世界中，这个概念肯定是有意义的——这个世界里没有心灵克隆术或者作用于大脑和思维的奇妙的复制粘贴操作。事实是，我们在平时的生活里或多或少都把“笛卡儿自我”视为一个理所当然的概念；它植入了我们的常识、我们的语言和我们的文化背景当中，根深蒂固、心照不宣、天衣无缝，而且跟时间流逝或者移动的事物会保持同一性这些概念一样，低调内敛，不见天光。可帕菲特关心的是，笛卡儿自我这种原生的概念在面对极端和史无前例的挑战时，还能不能站稳脚跟。作为一个审慎的思考者，他所做的工作跟爱因斯坦想象自己以光速或者接近光速的速度移动时所做的工作一样——他在把经典的概念推向极限的边界，而且同爱因斯坦一样，他也发现经典的世界观在与他们出生和成长大不相同的世界里并不总是有效的。

我是在金星，还是在火星？

帕菲特关于这个问题展开了100多页的沉思。他分析了很多思想实验，有些是他自己的创意，有些是其他同时代哲学家构思出来的，而他的分析永远都一针见血，清晰分明。我无意在此重复那些思想实验或他的分析，可我还是要总结一下他的结论。他的立场的核心实质是，人格同一性当被推至其边界时，便成为一种不确定的概念。在场景二这样的极端条件下，“他们中的哪一个才是我？”这个问题并没有一个有效的答案。

对于帕菲特这本书的很多读者而言，这是一个令人极不满意的答案，而且对于本书的很多读者而言，这个答案也难平众议。在地球上长大的我们所具有的直觉根本没有准备好应对这种非毁灭式的心灵传输的情境，所以我们鼓噪地想要一个简单直白的答案，可是我们的直觉又不知通过什么方式让我们感觉到，直截了当的答案根本就不存在。毕竟，我们还可以发明出场景三，创造一个如场景一中的那种毁灭式的心灵传输情境，但是可以把信号同时传送到金星和火星上的接受站。在这个情境中，帕菲特原来的身体和大脑被摧毁后，两个全新的帕菲特（都带着刮胡子留下的疤痕）差不多同时在两个星球上被组装出来。这回，关于二者的地位高下，似乎真的不存在任何有效的声明了（除非你争辩说，第一个完成的应该获得认领笛卡儿自我的荣幸，但是在这种情况下，我们只要设定他们两个完全同步地被组装起来，就可以把这条逃生路线轻易地堵死了）。

对于我们平常的、接地气的、怪圈642号式的头脑来说，这个情况非常明显，也十分简单：有一个帕菲特是假的。我们无法想象同时出现在两个地方，所以我们认为（把我们自己认同为大无畏的旅行者）：“我要么就是金星的那个，要么就是火星的那个，要么就哪个都不是。”可是这些答案里，没有一个能让我们经典的直觉满意。

帕菲特自己得出的答案其实更接近我在上一段里粗暴否决的想法：我们同时出现在了两个地方！我之所以说它更接近那个答案，而没有说它就是那个答案，是因为帕菲特的观点跟我在本书中提出的观点一样，都是在强调，那些在我们看来是非黑即白的东西，其实都是由不同的灰度构成的——只不过在平常的条件下，它们总是过于接近纯粹的非黑即白，以至于任何有关灰度的线索都藏而不见。其背后的缘由不仅来自明显的外部事实，即我们都拥有位于不同脑壳里的独立的物理上的大脑，还来自那张广泛的语言和文化传统网络，集体无意识坚持要我们相信，我们每个人都是精确的一个人（这是第18章中的“笼中鸟隐喻”，也是笛卡儿自我的概念），这在无形中阻止我们去想象任何一种灵魂的混合、重叠与共享。

我不能否认的是，在我们每个人的心底深处，还有一种绝对化的确定性，即我不可能同时出现在两个地方。在更早的章节中，我用很大的篇幅列举了这个观念的各种反例，而帕菲特也花费了极大的努力，给出了这种分散的同一性的其他证据来说明其可能性。事实上，他回避了“人格同一性”这个说法，更愿意用另外的术语来代替它，降低了唤起不可切分的“灵魂量子”（可类比于独一无二的出厂序列号或政府发行的身份证）这一意象的可能性。帕菲特更愿意使用的术语是“心理连续体”[注]，而我更愿意把他的意思称为“心理相似性”。换句话说，虽然帕菲特没有提出任何带有数学气息的说法，可他实际上却提出了一个抽象的在“人格空间”中不同人格之间[注]的“距离函数”（数学家会称之为“度量”）。

使用这种心灵对心灵的“度量”，我就会跟昨天的我非常“接近”，而跟两天前的我稍微远一点，以此类推。换句话说，虽然今天的侯世达和昨天的侯世达这两个个体之间有巨大的重合度，可他们并不是同一的。尽管如此，我们按照一般的标准（而且是下意识地）把他们视为同一个人，因为这么做是如此便利，如此自然，又如此简单。它让生活简单多了。这种惯例令我们能够给事物（不管是活物还是非活物）起上固定的

名字，日复一日地谈论它们，而无须更新我们的词典。不但如此，这个惯例从我们还是婴儿开始就扎根在我们的内部——根据皮亚杰提出的发展阶段，我们在这个时候开始明白，当一个球滚到一个盒子的后面时，虽然我们看不见它，可是它仍然存在，而且还可能在一两秒之后从盒子的另一侧再次出现！

帕菲特观念的本性

这些无意识的信念如此深刻地扎根于我们的世界观中，而且占据着很高的优先地位；打破这些信念，是一项极其鲁莽、且令人望而却步的任务，其微妙性与困难度堪比爱因斯坦在创造狭义相对论时完成的壮举(注)（经由纯粹的逻辑，颠覆我们关于时间本性的最根深蒂固的、最确切无疑的直觉）以及一整代以爱因斯坦为核心的杰出物理学家们开创量子力学的集体成就(注)（颠覆我们关于因果性和连续性之本性的最根深蒂固的、最确切无疑的直觉）。帕菲特提出的新观念是对于何为存在的颠覆性再感知，而且它在某种意义上是相当令人困扰的。但从另外的角度来看，它也是对思想的一次重大解放！帕菲特甚至花了一两页的篇幅，专门解释这个关于人类存在的激进的新观念如何让他得以解放，并深刻地改变了他对待自己的生活、死亡、爱人和其他人的态度。

《理与人》的第12章起了一个大胆的标题："我们的同一性为何不重要"。这一章中一系列发人深省的思考都配上了绝妙的充满挑衅性的标题。鉴于我对这本书如此仰慕，也很欣赏它的风格，我在此干脆就把这些章节的题目为你展示出来，希望借此可以引发你想要阅读那本书的兴致。这些标题是："分裂的头脑""用什么解释意识的统一性？""我分裂时发生了什么？""我分裂时重要的是什么？""为何没有能满足两个令人信服的要求的同一性准则？""维特根斯坦和佛陀""我实质上是我的大脑吗？"以及最后一个，"真的观念是可信的吗？"

尽管所有八个部分都富有洞见，我最歆慕的还是最后一节，因为在最后，帕菲特问自己是否真正相信他刚刚建造起来的宏伟的知识结构。这就好像爱因斯坦刚刚意识到自己的观念会把牛顿力学踩在脚下时却停下来问自己："我对于自己的思想方法真的有如此深厚的信仰吗？我能相信我刚刚得到的那个古怪而又反直觉的结论吗？我驳斥了一整套由在我之前的卓越物理学家们经过两三个世纪的时间小心翼翼地编织起来的自圆其说的环环相扣的观念网络，我是不是太过自大了呢？"

尽管爱因斯坦终其一生都是个极其谦逊的人，他给自己的答案（据我所知，他从来没有写过这样的自省文章）实际上就相当于："是的，我确实对我自己思想的正确性有一种奇怪的信仰。自然必须是这样的，不管其他人在我之前怎么说。我有幸获赠了朝自然的内在逻辑投以一瞥的机会，这一瞥比在我之前的任何人曾做到的都要更加深刻和准确。在这件事上，我的幸运无与伦比，虽然我不觉得这是我个人的功劳，但是我想要发表它，因为这样我才能跟其他人分享这一无价的景观。"

自信、谦虚和自我怀疑

帕菲特远比这谨慎多了。在我看来，他的结论跟爱因斯坦的发现一样激进（虽然我不太能想象关于人格同一性的难以言明的激进观念如何能催生不可思议的科技成果，而爱因斯坦的观念显然做到了这一点），但是他并没有像爱因斯坦一样，对自己的这些观念确信不疑。他对于自己的思想体系是有自信的，但不是绝对的自信。他并不认为如果自己站在这座大厦上，它将动摇并很快倒塌，可是他又坦承，也许是有这个可能的。让我们来听一听，关于这个话题，他自己是如何表述的：

> （心灵哲学家托马斯·内格尔）一度主张，即使还原论者的观点为真，对我们来说在心理上相信这一点也是不可能的。因而我将简短地回顾我在上文所给出的论证。然后我将询问我是否能够诚实

> 地主张相信我的诸结论。如果我能够做到，那么我将认为我不是独一无二的。至少会有某些人能够相信真理。
>
> （几页之后）……现在我已经回顾完了支持还原论者的观点的主要论证。我发现相信这个观点是不可能的吗？
>
> 我的想法是这样的。我能够在理智的或者反思的层面上相信这个观点。我折服于赞同这个观点的那些论证。但是我认为情况很可能是，在某个别的层面上，我将总是有疑问……
>
> 我猜想，回顾我的论证决不会整个地消除我的怀疑。在反思的或者理智的层面上，我会仍然坚信还原论者的观点为真。但是在某个较低的层面上，我会仍然倾向于认为在某个未来的人是我与某个未来的人是别人之间必定总有一个真实的差异。当我在摩天大楼顶层临窗观景的时候——类似的情况亦然，我知道我并不处在危险之中。但是从这个令人头晕目眩的高度向下观望，我感到害怕。如果我将要按下那个绿色按钮，我会有类似非理性的恐惧。
>
> ……要泰然地相信我的还原论者的结论有一定的难度。相信个人的同一性并不重要这一点是困难的。如果明天某个人将处于剧痛中，那这剧痛是不是会被我感觉到的问题很难相信是空洞的。而且承认如果我失去意识，那么“我将会死去吗？”这个问题便没有答案也是很困难的。

我必须要说，帕菲特愿意面对他的自我怀疑，并拿来与他的读者分享，我觉得这种表现极其可贵，而且带给人一种美妙的耳目一新之感。

把帕菲特变形为波拿巴

在上面引文的最后一段中，帕菲特提到了一个思想实验，这个实验有一部分是哲学家伯纳德·威廉斯（Bernard Williams）发明的，有一部分是他自己发明的（换句话说，它是由“威廉斯-帕菲特”发明的，也可

以叫“伯纳克·威菲特”）。在这个实验里，他即将接受一种特别的神经手术，而手术的性质是由一个数字参数决定的——也就是植入转换器的个数。这些转换器各自有什么作用？它们每一个都会把帕菲特的一种人格特征转变成一种不同的只属于拿破仑·波拿巴本人（我很快就会解释，这确实是“只属于他本人”）的人格特征。例如，一个转换器让帕菲特变得更加暴躁，而另一个转换器则让他不再排斥目睹杀人场面的想法，诸如此类。注意，在前面那句话里，我使用了“帕菲特”这个专有名词和代词“他”，后者想必是明确无疑地在指帕菲特。不管怎么说，这里的问题并不仅限于讨论这种用法是否合适。如果转换器越加越多，就会把帕菲特越来越多地转变成拿破仑，那么到了哪个阶段才会只有拿破仑存在——或者不如说，到了哪个阶段，这个缓慢变形的人才会完全成为拿破仑？

就像我在之前已经表明的那样，从帕菲特的视角出发，追问这种转换具体发生在变化线路上的哪一个点上是没有意义的，因为这里最重要的是心理连续性（即在我于前文不远处提出的那种大脑之间或人格的类数学空间内的接近性），而这是一个以不同灰度呈现的特质。它不是一个0或1，全或无的特性。一个人可以有一部分是德里克·帕菲特，而另一部分是拿破仑·波拿巴，而随着转换器越加越多，逐渐从一个向另一个滑动。而这不仅仅意味着，这个人变得越来越像拿破仑·波拿巴——这意味着这个人真的在慢慢变成波拿巴本人。

根据帕菲特的观点，拿破仑的笛卡儿自我并不是不可切分的，德里克·帕菲特也一样。这就好像在一根电线上装了一个滑块，通过把滑块在电线上滑动到任何一个想要的位置上，可以任意地把两个个体（其在语源学的意义上并非真正的“个体”，因为“个体”这个词本身就意味着“单一不可分”）融合或变形。结果就造出了一个杂合的人，距离两端有十分之一、三分之一、一半或者四分之三的距离——随便想调成什么比例都行，从德里克·帕菲特到德里拿·帕菲特，到德里拿·帕菲巴，到德破仑·帕拿巴，到拿破仑·波拿巴。

跟帕菲特不一样，大多数人都希望而且确信，在那个波谱上的每一个点上，对于“这个人是德里克·帕菲特吗？”这个问题都必定有一个明确的是或否的答案。当然，这是一个经典的观念——这个观念把帕菲特自己的笛卡儿自我视为理所当然的前提。所以，大多数人都陷入了一个尴尬的境地，他们不得不说，在这条电线的某个特别的点上，当滑块滑到此处时，突然之间毫无任何征兆，帕菲特的笛卡儿自我会噗的一声消失不见，被拿破仑·波拿巴的自我取代。就在刚才，我们面对的还是帕菲特，虽然他的人格多少有些改变，可是仍然是一个真切感受到德里克·帕菲特的感觉的德里克·帕菲特，可是现在，我们眼前的人突然之间变成了一个人格有所变化的拿破仑了，而他感受到的是拿破仑的感觉，而不再是帕菲特的感觉了！

侯世达的激进改编

我们在这里挑战的直觉是相当情绪化的，而且深深地根植于我们的文化和整体人生观之中。当我把自己放到这个场景中开始想象一个神经外科医生要通过植入一个又一个转换器来展开人格特征的转换时，这种直觉的冲突就变得格外激烈了。

例如，我最开始想象的是，当转换器1号装置完毕后，我对肖邦和巴赫的爱消失了，而对于他们的音乐形成了一种发自内心的厌恶，取而代之的是，在“我的”大脑里突然萌发了对贝多芬、巴托克、猫王和埃米纳姆的强烈崇拜。

接下来，我想象转换器2号使得我不再把每个周末（以及任何其他空闲时间）都花在设计双向图或努力完成一本有关一个怪圈的书上，转而选择把好几个小时的时间花在一个巨大的电视屏幕上观看职业橄榄球赛，并兴高采烈地朝啤酒广告里的大胸美女抛媚眼。

然后（转换器3号）我又想象自己的政治知识被倒转过来，包括我几十年间对性别歧视语言的长期抵制。现在，我要把“哥们儿”挂在嘴边了，而当有任何人表示不满时，我都会嘲弄她或他是“一只政治正确的猴子”（你可以想象，这只是我使用的言语中相对温和的一种）。

下一个转换器，让我放弃了自己一生秉持的素食主义倾向，转而染上了一种猎杀小鹿和其他野生动物的癖好——当然，这些动物的体格越大越好。所以，在转换器4号设置完毕后，我就*爱上*了用我生锈的来复枪射翻大象和犀牛的感觉！那简直是世界上最有趣的事了！而每一次，当这些高贵的野兽卑微地屈膝臣服在我的子弹之下时，我都会甩起胳膊，做一个“我棒极了”的姿势，就跟足球运动员射门得分之后的庆祝动作一样。

最后，无须多言，当转换器5号植入之后，我完全同意了约翰·塞尔的*中文房间实验*，而且我认为德里克·帕菲特关于人格同一性的观念完全就是一堆垃圾。哦，我忘了——我根本不可能想这些，因为我从来都不会思考有关哲学的东西！

你可能已经注意到了，当我在讨论转换器1号时，说到那个萌发了对贝多芬、巴托克、猫王和埃米纳姆的崇敬之情的大脑时，我在“我的”这个词上加了双引号。可是从那往后，我就不再麻烦加引号了，但是我可能应该都加上。说到底，我在前面几个段落里提到的每一件事都跟我认为的*核心“我”*截然相反。哪怕只让我放弃其中的一项特质，都会让我觉得：“那个人不再是我了。那*不可能*是我。那跟我的生命最深处的底里是不搭的。”

我们当然可以想象一些更温和的变化，比如在一个别样的人生里，我从来没有遇见普罗科菲耶夫的《第一小提琴协奏曲》。那可能是另一个版本的我，当然是一个没那么丰富的我，可是在现在的我看来，始终还是感觉像我。或者我们也可以想象，我仍然时不时地吃猪肉汉堡，但是会为此心存愧疚，或者在少见的场合下，我也会主动打开电视机看橄

榄球赛。这些灰度在道格的四周围成了一圈“可能的道格”的光晕，而位于中间的那个道格，正是我恰好成为的那个道格，这个道格的形成，取决于几十年以来在我身上发生过的数百万例偶然的事件，也有赖于刚好走进我生活当中的数百个特别的人（以及上百万个从未走入我生活当中的人，更别提还有更多的从未走入我生活当中的虚构的人物）。我们在正常情况下不会用这种灰度的形式来思考“我是谁/什么/如何”的问题，可我就是这么思考了，并在此吐露了一二。

关于“who”和关于“how”

我可能要顺便多说一嘴，我认为“who”这个词有时候被赋予了过高的潜在力量，跟人称代词“他”和“她”的情况差不多（你也许还能记得，我曾在第1章里跟凯莉展开过一次关于把人称代词用在动物身上的简短讨论）。20世纪80年代，帕梅拉·麦考达克（Pamela McCorduck）写了一本关于人工智能历史的书，书名既激动人心，又别具一格，叫《会思考的机器》（*Machines Who Think*）。标题中的“who”这个词创造了一种全新的意象，同我们与标准机器如膝跳反射一般联系起来的印象——如开罐器、电冰箱、打字机甚至计算机——根本不同；这个词暗示着，至少存在某些机器，“在那里”有某个人，或者用托马斯·内格尔的话说，“身为那台机器有某种特别的感觉”（顺便一提，这句话很难翻译成英语之外的其他语言：there is something it is like to be that machine）。这再一次内含暗示，在一系列假设性的“在思考的机器”[注]（这样的机器仅仅在思考，但是没有内在生命）和一系列不同的“会思考的机器”（Machines who think，这些机器会有一个内在生命，而且每一台都将成为特定的那一个）之间，有一条清晰分明、非黑即白的二分线。我常常觉得，归根结底，当我在思考我最亲密的朋友是谁的时候，总是会归结到他们的“how”（如何）之上——他们如何微笑、如何说话、如何大笑、如何倾听、如何承受、如何分享等等。我告诉我自己，每一位朋

友的最内在的实质是由成千上万这样的“如何”构成的，而这些“如何”的集合，正是“这个人是谁？”的答案——完整的答案。

这看起来好像纯粹是一个第三人称的外部视角，而且排除甚至完全否定了第一人称体验。这看起来故意削减了“我”，甚至导致了“我”的消散。可是我不这么认为，因为我觉得，即便对于它自己而言，“我”也不过就是这么一回事而已。问题在于，一个“我”总是很擅长说服自己，相信自己远不止于此——事实上，这正是“我”这个词的全部使命！“我”是维持这场骗局的既得利益者（尽管它也是自己骗局的受害者）。

要么翻倍，要么输光

过了这么久，我们终于又回到了场景三中金星与火星对峙的困境。我已经告诉过你，帕菲特多少有点避重就轻地绕过了这个问题，只是简单地否定了笛卡儿自我的存在，从而宣布这个问题没有任何有意义的答案。但是在他的书里，他还相当频繁地提到过他所谓的“双重生存”，意思是他实质上同时出现在了两个地方。他不止一次地写到，双重生存很难跟死亡画上等号（后者指的是无一生存），而且数字2也不能跟数字1混为一谈！所以他到底在说什么呢？他是在说这个问题没有答案，还是在说事实上他被“翻倍”了，于是现在有了两个德里克·帕菲特呢？

我很难搞清楚这一点，因为我觉得他这两种说法出现的次数都足够多了，任凭人们怎么说都有道理。但是我对这个问题到底怎么看呢？我想我最后还是要站在“两个我”的那一边。乍听上去，我几乎好像是在支持笛卡儿自我的学说，只需要想象那个蛋被克隆出来，然后两个一模一样的蛋就同时存在了，一个在金星上，另一个在火星上。可是，怪圈642号随后开始高呼：“哪一个才是我呢？”听起来，仿佛我根本就没有回答这个问题，或者说，仿佛我想让我的蛋在火星上出现，又要在金星上吃掉它。

为了不再自相矛盾，找回一点逻辑连贯的感觉，我不得不再次返回那段对话中怪圈641号的主旨，即“我”这个概念从根本上来说都只是一种幻觉。让我们把场景三应用在我的身上，而不再是帕菲特。心灵传输在金星和火星上制造了两个全新的我，而地球上没有留下一个副本。在这种情况下，每一个新的大脑——火星上的那个和金星上的那个——都确信它就是*我*。它给人的感觉像是它一直都感觉自己是我。当有人敲打我的膝盖，让我的小腿弹起来的时候，两个大脑中都会自动地涌出那个老调的冲动，想要开口说：“我在*这里*，而不在*那里*。”但是不管有没有膝跳反射，事情的真相是，名为“我”的东西压根儿就不存在——没有硬玻璃球，也没有笛卡儿蛋壳保护下的珍贵蛋黄，真正存在的只有倾向、癖好、习惯，包括言语上的那些[注]。说到底，当两个侯世达同时说“*这里*的这个才是我”的时候，我们都不得不两个都信，我们相信这两个侯世达的程度至少也跟我们相信正在书房里打出这些文字并用印刷文字跟你说“这里的这个才是我”的侯世达的程度一样。这种说法以及认为它是真相的坚持，只不过是一种倾向、一种癖好、一种习惯而已——事实上，就是一种膝跳反射——尽管它从表面上看起来绝不止如此，可事实上真的就仅此而已。

归根结底，“我”就是一种幻觉，可与此同时，充满悖论色彩的是，它又是我们所拥有的最宝贵的东西。正如丹尼尔·丹尼特在《意识的解释》（*Consciousness Explained*）一书中所指出，一个“我”有一点儿像一张纸币——它*感觉起来*好像价值不菲，但说到底，只不过是一种社会契约，一种我们共同默认的错觉，从来没人会问我们这回事儿；它虽然是幻觉，却支撑着我们的整个经济。可是，纸币到底不过是一张纸而已，没有任何内在的价值。

滚动的列车

从第15章到第18章，我一直在主张，我们每一个人都是分散的，而且尽管与我们通常的直觉相悖，我们每个人都至少有一部分居于四散在这个星球各地的不同大脑里。这个观点等同于在说，一个人可以同时出现在两个地方；虽然在一开始，我们会膝跳反射般地拒绝这种听起来很疯狂的想法。如果同时在两个或者更多地方出现听上去仿佛是不可理喻的，那么不妨想想时间和空间的变换。也就是说，可以想见，让你想象自己在明天或者后天仍然存在绝非难事。而在这些未来的人当中，哪一个才真的是你呢？怎么可能有两个不同的你同时存在，还都在认领你的名字？“啊，”你回答说，“但是我很快就能到那儿了，就像一部列车经停不同的车站一样。”可是这个回答回避了问题的实质。为什么那还会是同一部列车呢？在这个过程中，可能有不同的乘客上下车，它可能更换了一两个车厢，也许甚至连发动机都换过了。它只不过是名叫“列车641号”而已，而这正是它为什么还是“同一部列车”的原因。它是一种语言学上约定俗成的惯例，而且是一种很好的惯例。它是我们所在的经典世界里十分自然的一种惯例。

如果列车641号从米兰出发向东行驶，在维罗纳分成两列，一列开往波尔扎诺，一列继续向东，开往威尼斯，那么我们也许就不能再把其中任何一列称为“列车641号”了，而是要给它们分配不同的数字。但是我们也可以叫它们“列车641甲”和“列车641乙”，或者甚至干脆仍然继续叫它们“列车641号”。毕竟，还有可能发生下面这样的事：向北行驶的一半列车在抵达波尔扎诺之后突然转向往东行驶，而向东行驶的那一半也在抵达威尼斯之后突然转向北上，两个部分在贝卢诺汇合，重新连在一起，继续它们的——或者说它的——行程，向终点乌迪内出发。

你也许会反对说，列车在整件事上没有内在的角度——“641”只不过是一个第三人称的标签，而不是第一人称的视角。我只能说，这个观点很诱人，但是站不住脚。会滚动的列车（trains who roll）和在滚动的列车（trains that roll）其实是一样东西；至少，如果它们拥有足够丰富的表征系统，能够包含并表征自身，那么事实就是如此。今天的绝大多

数列车都没有（事实上，没有任何列车有这种表征系统），所以我们通常不会把代词“who”用在它们身上。但是也许某一天它们就有了，而我们也会这么用。不论如何，从一个代词向另一个代词的转变并不是骤然而分明的一刀切；它是逐渐变化的，就像人们在思想的复杂度上不断成长，对笛卡儿自我的信仰也会逐渐地褪色而消失。

灵魂日冕之光

整章读下来，你可能会产生这样的印象：整个一章的论述都是以一个科幻场景作为前提的，完全没有触及我们思考真实人类所处的现实世界的方式，罔顾他们现实的生与死。可是我相信，这只是一场误会。

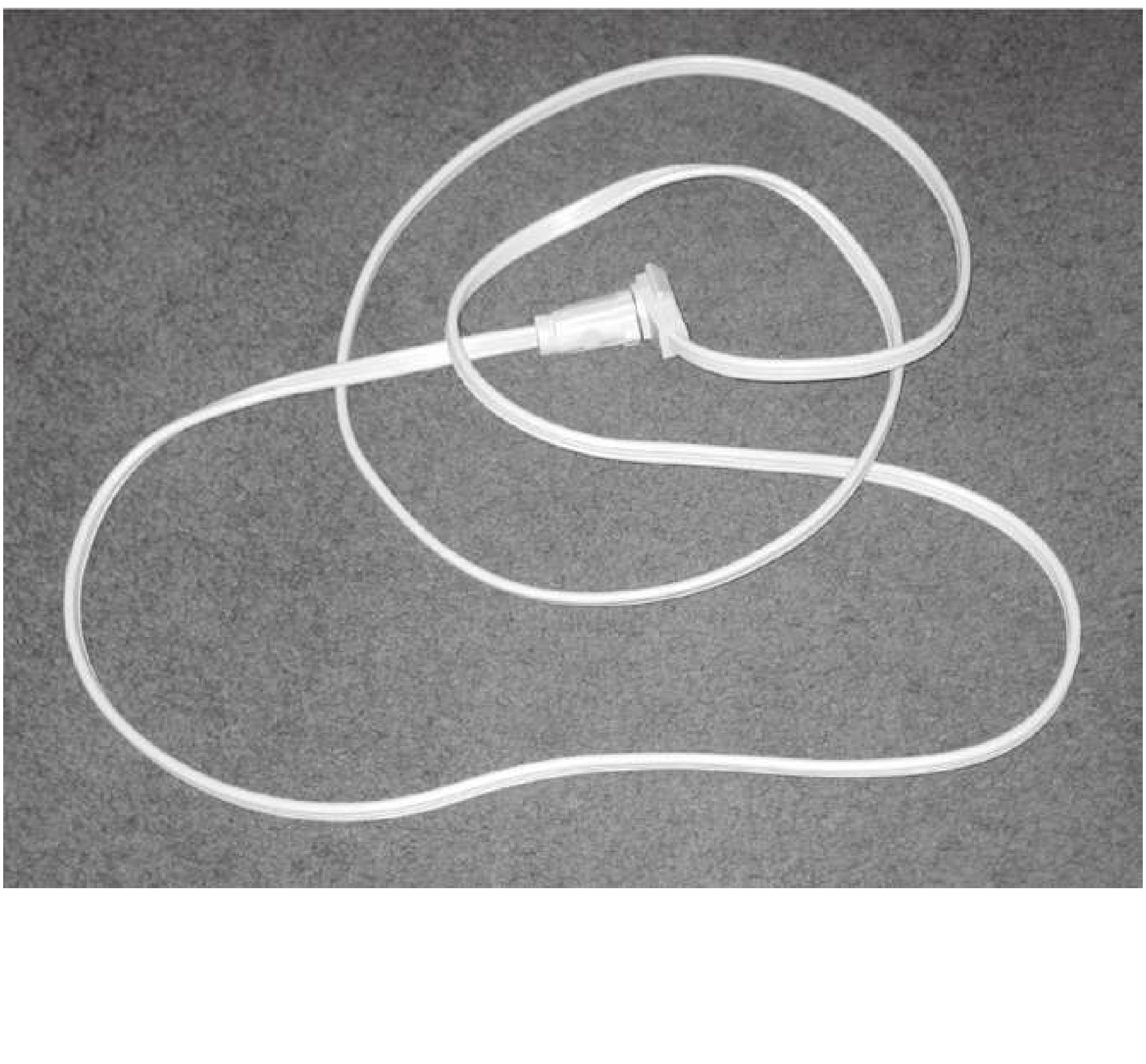
我有一位关系很近的朋友，他的老父吉姆患上了阿尔茨海默病。我的朋友在几年的时间里悲伤地眼看着他的父亲一点一点地与现实脱节，而那些现实就在几年前还是构成他内在生命的绝对基石，是完全可靠的结实陆地。他记不住他的家庭住址了；他丢失了之前有关尘世之物的认识，比如信用卡；而且他也不太确定他的孩子们是谁了，虽然他们看起来还是模模糊糊地很眼熟。这一切都变得越来越黯淡，而且不会再亮起来了。

也许吉姆会忘掉他自己的名字，忘记他生长的地方，忘记他喜欢吃什么，以及更多。他正朝着里根总统在生命最后几年的低亨内克岁月里的所在之处逼近，步入那团可怕、浓厚、包围了一切的迷雾。可是，吉姆的某些东西还十分强壮地活着——活在其他的大脑里，而这都多亏了人类的爱意。他那亲切随和的幽默感、他从驾车在大草原的开放旷野上任性飞驰中收获的无尽欢乐、他的理想、他的慷慨、他的单纯、他的希望和梦想——以及（不管有多少价值）他对信用卡的认识。所有这些东西都在很多人内部的不同层面上存活了下来。这些人通过几年乃至几十年来与他亲密的交往，构成了他的“灵魂日冕”（soular corona）(注)——

他的妻子、他的三个孩子，还有他很多很多的朋友。

不等吉姆的身体物理死亡，他的灵魂就会变得相当模糊而黯淡，跟不存在没什么区别了——灵魂的日食遮天蔽日——可尽管日食虽不可避免，他的灵魂犹在，其低分辨率的局部副本散布在这个星球的各地。吉姆的第一人称视角会在其他的大脑中时不时地闪现出来。他还会存在，一会儿在这里，一会儿在那里，虽然是以一种极端稀释的形式。吉姆将在哪里呢？不能说差不多到处都是，这一点必须承认，但是在某种程度上，他可以同时以不同的程度出现在很多地方。虽然光芒已经极其削减，可是但凡他的灵魂日冕还在，他就还在。

这很悲伤，可是也很美好。不论如何，这是我们仅存的慰藉。



1. Dennett introduced his term “intuition pump”, I believe, in the Reflections that he wrote on John Searle’s “Chinese room” thought experiment in Chapter 22 of [Hofstadter and Dennett].
2. See [Nozick] for a lengthy treatment of the closely related concept of “closest continuer”.
3. 或者说在不同的大脑之间；只不过，这种“距离计算”需要在何种结构层级上描述大脑，并没有明确的说明，而且也很难想象那个层面可能是什么。
4. See [Hoffmann].
5. See [Pais 1986], [Pais 1991], and [Pullman].
6. 在这里，将“Machines that think”翻译为“在思考的机器”，以区别与“Machines who think”（会思考的机器）。——译者注
7. See the Prologue for my first inklings of this viewpoint. See also my Achilles–Tortoise dialogue entitled “A Conversation with Einstein’s Brain”, which is Chapter 26 in [Hofstadter and Dennett], for more evolved ideas on it.
8. soular corona（灵魂日冕）和solar corona（日冕）是一个文字游戏中的类比，可惜中文难以体现原文的用意。——译者注

第22章
同僵尸与二元论舞起一支探戈

学究式的语义学？

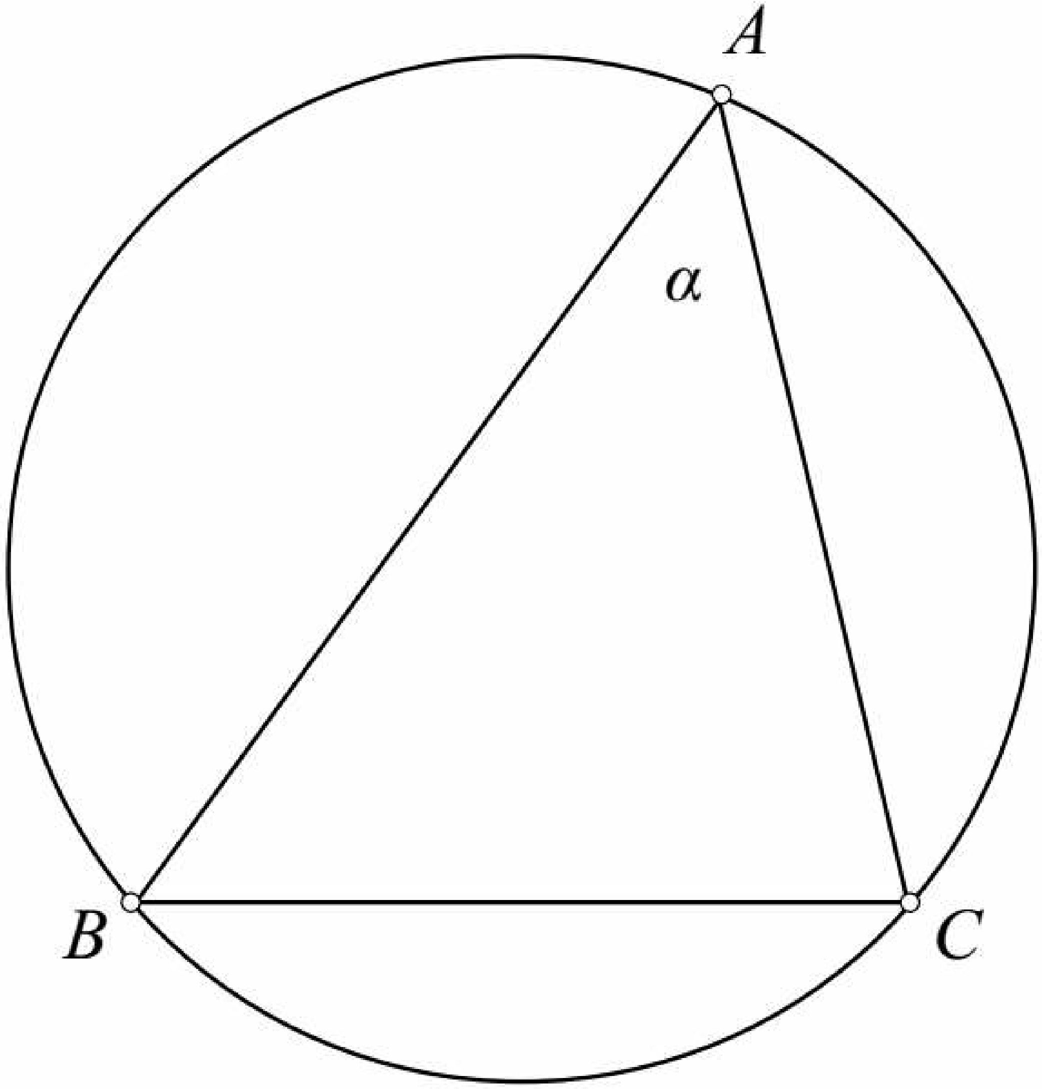
我们假想了一台思考机器，然后争论在未来某一天用哪一个关系代词来指称它才合适，到底应该是“who”（谁）还是“which”（哪个）。毫无疑问，有些人会把这番争论的实质当作一次学究式的在语义学上的吹毛求疵；可是对另外一些人而言，这场争论却引发了生死攸关的议题。没错，这个议题在本质上是一个语义学问题，因为它关乎用何种言语标签来标记某种前所未闻之物，但是鉴于范畴的分配直接贯通了思考的核心，它们也决定了我们看待世界上每样东西的态度，包括生与死的大事。正因如此，我觉得这个代词问题即便“仅仅是语义学上的”议题，对于我们关于自己是谁或者是什么的认识而言，仍然很重要。

澳大利亚著名的心灵哲学家大卫·查尔默斯（David Chalmers）曾经是我的博士研究生，现在是我的挚友。他投入了很多年的时间，全力声援那个充满煽动性的观念，即“machines that think”和“machines who think”有可能同时存在。对我来说，这两种类型的机器可以共存的观念毫无意义可言，因为正如我在第19章中所宣称的那样，“思考”这个词表示的是在一个闹壳（或者类似的其他场地）中的符号之舞，而这也是“意识”这个词的所指。既然有意识就意味着有权使用“who”这个代词（当然还包括“我”和其他代词），那么思考也就具有同样的效力——在我看来，问题至此就已经解决了。换句话说，“machines that think”是一个自相矛盾的短语，因为它的关系代词用的不对；而如果有一天，真的

出现了思考的机器，那么按照定义，它们只能是“machines who think”。

两台机器

大卫·查尔默斯用一种前所未见的新方式探讨了这个问题[注]。他想象出一个世界，在那里有两台一模一样的机器，它们的每根螺丝和晶体管，乃至每个原子和夸克都是一样的。它们都放在巴基斯塔尼亚大学的意识与认知功效研究中心[注]，并排摆放在641号房间的老橡木桌子上，执行着完全一致的任务。出于具体的考虑，让我们假设，两台机器都在奋力地用一种非正式的几何顿悟而非正式的代数运算来证明欧式几何中那个简单而又令人惊奇的“弦角定理”，根据这一定理，如果一个点（下图中的*A*）沿着一个圆形的弧线移动，那么这个点在移动过程中所“看到”的那个由它与一条固定的弦（BC）围成的角度（α）将是恒定不变的。



我之所以选择这个初级却又优美简洁的定理，是因为它是我和大卫在很多年前在一起愉快讨论过的一个定理，而他当时的某些评论给我带来的启发，切切实实地改变了我的人生。事实上，正是那个命运般的岔路口，让我得以想象出6号转换器，而那个转换器一旦生效，将把所有关于这个定理的知识从我的大脑中抽走，包括后来因我对这个定理的仔细思考而燃起的所有几何学热情……

正如我刚刚所说，这两台完全一模一样的机器在原子钟的精准定时

下同时启动了这项工作，并朝着任务的解决方向以精确一致的步调前进，我们可以说，它们精确地模拟了当大卫·查尔默斯首次在视觉上找到让自己恍然大悟的证明时他大脑中发生的事情。在此，对于我们而言，这两台机器中程序运行的细节并不重要，重要的是机器Q（Q代表“qualia”，即可感受的特质）真的在感觉某种东西，而机器Z（Z和“死”谐音）则感觉不到任何东西。就是从这一点开始，大卫的观念让我渐渐无法理解了。

我不得不承认，为了让读者更容易想象，我把大卫讲的故事进行了轻微的篡改。我把这两台机器并排摆在了CRCC（意识与认知功效研究中心）641号房间的老橡木桌上，而大卫从没提到过这一点。事实上，他大概会说出下面这样的话，来提出抗议：“假设在同一张老橡木桌子上有两台运行一模一样的过程的一模一样的机器，其中一个能感觉到某些东西，而另一个则不能，这根本就是狗屁不通。彻底违反了宇宙的法则！”

我完全接受他的反对，而且对自己歪曲大卫故事的罪行供认不讳。为了减轻我的罪孽，让我的故事变回他的版本，我首先要把其中一台机器从641号房间的老橡木桌上搬走。不管我们之前是怎么命名的，都让我们把那个留下来的机器称为“机器Q”。下面（跟着大卫的思路），我们迈出了出乎意料的一步：我们想象出了一个不同的而又同构（即“既独立又难分彼此”）的宇宙。我们把第一个宇宙命名为“宇宙Q”，而叫新的那个“宇宙Z”。两个宇宙遵守完全一致的物理法则，而且在每个宇宙中，人们只需要了解物理法则，就足以预测任何给定的初始粒子接下来将会发生之事。

当我说这两个宇宙难分彼此时，万般后果之一，就是宇宙Z跟宇宙Q一样，也有一个银河系，里面有一颗名为“太阳”的恒星，引领了一个由9颗行星组成的恒星星系[注]，其中第三颗行星的名字是“地球”，而在宇宙Z的地球上，有一所巴基斯塔尼亚大学，开设了一个意识与认知功

效研究中心，我们的老朋友，641号房间，就在这个中心里面。房间里甚至有“同一张”老橡木桌子，然后请看，“同一台机器”也放在上面。你肯定看到它，不是吗？但是，既然这台机器在宇宙Z里，我们就要叫它“机器Z”，这样我们才能用名字来区分这两台坐落在两种难分彼此的环境之中的难分彼此的机器。

当然，这样一来，我们就无法在“同一时刻”开启机器Q和机器Z了，因为他们属于不同的宇宙，各有独立的时间线；可是幸运的是，这两个宇宙具有完全一致的相同物理法则，所以同步化根本就没有必要。我们只要开动它们，让它们各干各的就行了。鉴于它们都遵循同样的物理法则，而物理学就足以决定精细到一分一毫的所有行为了，所以它们还是跟以前一样，在做一模一样的事。话说回来，你还指望有什么不同的结果吗？古怪的是，虽然两台机器在做一模一样的事，直到夸克乃至更小的层级上都没有差异，但是，机器Q享有对它所做之事的感觉，而机器Z则没有。机器Q为工作成果而感到狂喜，而机器Z毫无感觉。一点儿都没有。

“那怎么可能呢？”你可能要问了。我也要问同样的问题，而且跟你一样困惑不解。可是大卫却兴高采烈地解释说：“噢，这是因为，机器Q存在的那个世界里，除了物理法则之外，还有某种额外的东西，让某些特定种类的物理过程可以伴有感觉的发生。即使这些感觉没有对任何物理事物产生影响，也不可能有什么作用，它们仍然是真实的，而且它们真的就在那里。”

换句话说，虽然宇宙Q和宇宙Z中的物理法则是完全一致的，可是在宇宙Z中，却找不到感觉——只有空洞的动作。机器Z口中说着跟机器Q一模一样的话。它声称对自己找到的证明感到狂喜（跟机器Q一模一样），而且滔滔不绝地讲述它在这个证明中发现的美妙之处（跟机器Q一模一样）——可是事实上，它没有感觉到任何东西。它说的话都是中空的言语躯壳。

两个大卫

令宇宙Q和宇宙Z如此不同的那种额外成分是什么呢？大卫没有说，但是他告诉我们，那就是意识的材料——我给它起名叫“心灵活力”（élan mental）。如果你生在一个具备这种要素的宇宙里，那么你就很幸运了，可是如果你生在一个没有心灵活力的宇宙里可就惨了，因为那里没有你，没有我，没有谁，也没有你中的我（或者他或她）——而只有它。尽管有这种巨大的差异存在，两个宇宙中所有的客观现象都是一模一样的。所以，两个宇宙中都有马克斯兄弟（Marx Brothers）的电影，而当宇宙Z中的Z人们在观看《歌声俪影》（*A Night at the Opera*）时，他们会跟宇宙Q中的Q人们在观看《歌声俪影》时发出完全一样的大笑[注]。

最美妙的反讽是，不但宇宙Q中（*我们*生活的宇宙）有一个大卫·查尔默斯，宇宙Z中也有一个大卫·查尔默斯，它周游世界各地，发表学术演说，讲解为什么在它出生的宇宙里存在感觉，而在另一个同构的宇宙中却*不存在*感觉，而它那个不幸的“僵尸孪生兄弟”就出生在那个宇宙里。这里的反讽之处当然是，宇宙Z的大卫·查尔默斯满口谎言，可是它丝毫没有意识到自己在说谎。虽然它*相信*它是有意识的，可真相是，它并没有。令人悲哀的是，这个大卫是意识*幻觉*的一位天真无知的受害者。这种幻觉中的意识，只不过是深深扎根在它的大脑里的一个怪圈所带来的微不足道的副产品，而它在宇宙Q中的同构对应物，口中也用同样的语调说着同样的话，却是在陈述真理，因为*他*真的*是*有意识的！为什么？因为他不仅在他的大脑里有一个怪圈，而且这个幸运的家伙还生活在一个具有心灵活力的宇宙里。

请千万不要觉得我是在取笑我的朋友大卫·查尔默斯，因为大卫*真的*在周游世界，访问世界各地的哲学系，发表客座演讲，兴高采烈地描述他的“僵尸孪生兄弟”，欢欣愉悦地笑话那个孪生兄弟无助的幻觉，因为那个僵尸孪生兄弟也在一字不差地发表同样的演讲，说着同样的笑

话，相信它说出的每一个字，可是感觉不到任何东西。大卫是一位洞见非凡的思想者，而且他跟我一样清醒地意识到自己区分宇宙Q和宇宙Z、机器Q、机器Z以及他自己和他所称的僵尸孪生兄弟的做法，在表面上看起来是很疯狂的；但是，尽管我觉得所有这些言论都愚蠢到难以接受，大卫却深信这种区分虽然乍看上去是不讲道理的，但是宇宙Q中那种神秘的、非物理性的、缺乏因果性的额外成分，即心灵活力——怪圈641号和怪圈642号讨论过的“感素”概念的近亲——是意识本性的关键；正是因为这一关键发现的缺失，意识的本性才显得难解莫测。

关于一个人有可能是僵尸的萦绕不散的担忧

最近，不少心灵哲学家都跟大卫一样，跟风迷上了“僵尸”这个概念。（实际上，这更像是那种“让我们爱恨交加的概念”。）它似乎起源于加勒比海一带的巫毒仪式，由此传播到恐怖电影里，接着又进入了文学世界。上网搜索一下，你马上就能找到所有你想了解的信息，其中绝大多数都很有趣。

基本上，僵尸就是一种没有意识的类人，他们的行动——哎哟不对，我的意思是，“它们的行动”——表现得好像有意识一样。僵尸没有内在，虽然从外部看，人们可能认为是有的。我得承认，我偶尔会撞见某些目光呆滞的人，他们的眼神让我产生一种毛骨悚然的感觉，觉得那双眼睛的背后仿佛没有任何人存在。当然，我不会把这种印象当真。可是对于很多哲学家而言，那双空洞而无神的眼睛意象已经变成一种范式化的恐惧。如今不乏心灵哲学家发现，僵尸这一概念不仅令人深恶痛绝，而且事实上还有令人束手无策的内部连贯性。这些哲学家深受僵尸阴魂的困扰，于是主动担起一项神圣的使命，想要展现出我们的世界不是冰冷而空洞的宇宙Z，而是温暖而暧昧的宇宙Q。

现在你可能要说，这本书全篇都在把人类视作冰冷的、目光无神的

僵尸，因为它把“我”的本来面目设定为一种错觉、一种思维的障眼法、一套大脑欺骗自己的把戏和由幻觉幻想出来的一种幻觉。这就意味着，我们都是无意识的，但是我们都相信我们是有意识的，而且我们都表现得有意识。好吧，没问题。我同意，如此界定我的观点，还算公正公平。但是，那群有僵尸恐惧症的哲学家都想让我们的内部存在比上述的更加丰富。他们声称自己可以毫不费力地构想出一个冷若冰霜的宇宙，其中的居民都是噩梦一般空洞无魂的僵尸，可是那个宇宙无法以任何客观的方式与我们自己的宇宙进行区分；与此同时，他们又坚称，这不是我们生活在其中的那个宇宙的样子。根据他们的说法，我们人类不仅仅是表现得有意识或者声称自己有意识；我们真的就是有意识的，而且那完全是另外一回事儿了。因此，侯世达和帕菲特是错误的，而查尔默斯是正确的。

好吧，我觉得，丹尼尔·丹尼特对这些哲学家的批评正中要害(注)。丹指出，不管这些哲学家如何庄严地许诺，他们其实都没有构想出一个跟我们的世界一模一样而只是由僵尸占领的世界。他们甚至看起来都没有非常努力地去尝试过想象这件事。他们跟怪圈642号一样，当他想象一个怪圈看到一朵紫花时会说出什么的时候，他选择了一个去人性化的动词“嗡嗡叫”来形容它说话的方式，而且把它的语音比作遭人恨的电话菜单树里机械录制的声音。怪圈642号对于怪圈有一种刻板印象，认定它是没有灵魂的，这种偏见践踏了完全自然而正常的人类行为的画面。同理，那些畏惧僵尸的哲学家之所以害怕僵尸，是因为他们害怕那种机械化的嗡嗡叫，害怕那种呆滞无神的目光，还有那必然弥漫在仅仅由僵尸居住的世界里的残酷冷漠——可是就在不久之前，他们刚刚画押认可这样一个世界跟我们的世界是难以区分的。

意识不是一扇电动天窗

在有关意识的争论中，最常出现的问题之一大致是："到底是意识的哪一点在帮助我们生存下来？我们难道不能保留所有的认知工具，但作为感觉不到任何东西也没有任何体验的机器而存活吗？"在我看来，这个问题本质上是在问："为什么意识是在大脑达到某个特定复杂程度之后加上来的？为什么意识的加入，像是某种赠品？如果拥有意识对进化有额外的好处，那么这种好处到底是什么？"

提出这个问题的一个默认前提是，没有意识的大脑，也可以达到任意的复杂程度。这相当于认可了在641号房间里那张老橡木桌子上并排摆放的两台机器Q和Z之间的可区分性，虽然它们在执行同样的操作，但是其中一个是带着感觉在做事，而另一个则没有带着感觉。它假设意识是某种可以添加的"额外特质"，某些模型可能拥有它，也可能不具备它，哪怕最精致酷炫的模型也不例外，就像一台豪车可以配置一台DVD播放器或者一个电动天窗，也可以不装配这些设备。

但是，意识并不是一扇电动天窗（你可以随便引用我这句话[注]）。意识不是一个可选的特质，人们不能枉顾大脑的构造而任意添加。你不能选购一辆配有双缸发动机的汽车，然后又跟卖车的人说，"另外，请再给我来一套赛车马力®。"（当然，谁也不能阻止你下这个订单，可是我劝你千万别心心念念地盼望着到货。）就算有人买了一辆火辣的16缸发动机汽车，也没有道理去问人家："不好意思，我还要砸多少钱才能买到赛车马力®？"

这个愚蠢的"赛车马力®"是我生造的一个概念，实际上它就是马力等级的连续波谱的顶端，而这种马力等级是发动机从它们的设计中自动获得的结果。同理，意识也只不过是自我感知层级的连续波谱的顶端而已，是大脑从它们的设计中自动获得的结果。你和我的大脑都是100或更高亨内克的豪华赛车大脑，它们拥有大量的自我感知，因而也拥有大量的意识；像卷起的橡皮筋一样的初等大脑，例如蚊子的大脑，根本没有意识；最后，那些中等的大脑（两岁的孩子、宠物猫或小狗），亨内

克数不大，所以它们拥有的意识也只有一点点。

意识不是当一个人拥有一个100亨内克的大脑之后外加的一个选项，它是系统的范畴库足够精密复杂这个事实所带来的必然后果，将不可避免地呈展出来。哥德尔的怪圈，是在足够强大的数论形式系统中自动浮现出来的；同样，自我的怪圈，也将在任何一个足够精密的范畴库中自动生成。而一旦你得到了自我，你也就拥有了意识。并不需要什么心灵活力。

叶哲学

相信意识来自某种超越物理法则之上的东西的哲学家都是二元论者。他们相信我们居住在一个类似于魔幻现实主义的世界里，那里有两种类型的实体：具有心灵活力的魔法实体和缺乏心灵活力的普通实体。更具体地说，魔法实体拥有非物理性的灵魂，也就是说，这种实体中不多不少地灌注了一个“意识团”（团是心灵活力的标准单位），而普通实体没有这样的团。（大卫·查尔默斯相信的是两种类型的宇宙，而不是一个单独宇宙中两种不同类型的实体，但是在我看来，这都是相似的二分法，因为我们完全可以把不同的宇宙也看作是一个更大的“巨型宇宙”内部的实体。）现在，亲爱的读者，我想要十分清楚地确定，关于这种魔法实体与普通实体的二分法，你跟我是否意见一致。因此，为了让这一点清楚到最大的程度，我要在下面进行一番“温柔的”演绎。

想象有一所名为“叶哲学”（Liphosophy）的哲学学院，门徒都叫“叶哲学家”，他们相信有一种难以捕捉——实际上是难以察觉——而又极为重要的非物理性性质的存在，名为“自我之叶”（Leafpilishness）。他们还相信，在我们的宇宙中，有某种特别的实体，充满了这种快乐的品质。接下来，不怎么令人惊讶的是，这些上天保佑的实体，就是你和我更愿意叫成“叶桩”（leaf piles）的东西（先不管这个短语表示的意思有

多么模糊不清）。如果你我瞥见一个这样的东西，而且正好心情到位，那么我们可能会大声叫道："嘿，你知道吗？那是一个叶桩！"我猜，这样一次热情的爆发对你我而言都是极限了。我们都不太可能会在这个场景里驻足徘徊。

但是对一个叶哲学家来说，这却会引出更多的思考："啊哈，这又是那些稀有实体中的一个，它们被注满了一团自我之叶。这种神秘的、非物质性的、超越世俗而又十分真实的气息，绝不会附着在干草垛、草纸堆或者薯条上，而只依附在叶桩上！如果没有自我之叶，那么一个叶桩就只不过是一堆杂乱的树叶残渣；而多亏了有自我之叶的存在，这些杂色的叶子堆才变成有我之叶！又因为每一团自我之叶都与众不同，所以地球上的每个叶桩也就获得了一个完全独特的同一性！自我之叶是一种多么神奇而深邃的现象啊！"

亲爱的读者，不管你对于意识持有什么样的看法，我猜你都会对叶哲学的信条感到大惑不解。"那个疯疯癫癫的自我之叶到底是怎么一回事儿？[注]染上这种隐形的、无法察觉的气息之后，接下来会发生什么？"要是你心里没有想到这些问题，那才是怪事呢！你可能还会想，"决定这个物理世界中的哪些实体会接收这一团团自我之叶的究竟是谁？或者是什么？"

这些思索也许会将你引向另外的一些难题，比如：准确地说，构成一个叶桩的是什么？要构成一个叶桩，需要多少片叶子，对它们的尺寸有什么要求？哪些叶子是属于这个叶桩的，哪些不属于？"属于"一个指定叶桩，是一件非黑即白的事吗？叶子间的空气怎么说呢？叶子上的尘土怎么算呢？要是叶子太干，有一些（或者有一半，或者大部分）被揉成了微小的碎片，又将如何？两个邻近的叶桩会不会共享位于它们之间的一些叶子呢？一个叶桩的边界是不是永远都百分之百的清晰可辨呢？简言之，大自然是如何凭借一种非黑即白的绝对方式来判明哪些东西有资格接收这一团团的自我之叶呢？

如果你有心情展开更进一步的哲学思考，那么你可能还会问：如果出于某种离奇的意外或者古怪的错误，一团自我之叶附在了一个有一只蚂蚁在里面爬行的叶桩（也就是说，它附在了一个由叶子加蚂蚁组成的合成体上），又会发生什么？或者如果它只附着在一个叶桩的上面三分之二的部分呢？或者如果它附着在一个海草桩上了呢？或者如果是一个孩子在海滩上堆起来的摇摇欲坠的沙堡呢？要是旧金山动物园呢？要是仙女座星系呢？或者我下周的牙医预约门诊？如果两团自我之叶不小心同时附在了一堆叶桩上，又会怎么样？（或者没有任何附着，造出一个“僵尸”叶桩呢？）有哪些可怕或美妙的结局会随之而来呢？

我猜，我的读者不会把一个叶哲学家的话当真，相信自我之叶是宇宙的一个神秘的成分，认为它超越了物理法则，凡是具备自我之叶的东西都与宇宙中所有其他东西有内在的不同，而且每一个叶桩都有一个独特的同一性——这个同一性并非来自它独特的内部构成，而是得益于不知道从哪里生长出来的一团特别的自我之叶。我希望你愿意跟我一起大声说出：“叶哲学就是一堆混乱的信仰！”并不把它放在眼里。

意识：一种用楷体呈现的精华物

不说叶哲学家了。下面让我们来谈谈那些正经的哲学家，他们把意识视为世界的一个难以捉摸——实际上是难以察觉——而又极其重要的非物理性方面。为了把这个意识概念跟我在整本书中一直在讲的意识概念做出区分，我下面要把意识这个词变成楷体，把这个意识称为“意识”。每当你看到意识这个词变成楷体的时候，只要想想那个名为心灵活力的非物理性精华物，或者跟赛车马力®或自我之叶进行一下类比，就能把握个八九不离十了。

在这一点上，我不得不承认，我对于这个用楷体呈现的精华物的想象相当贫乏。为了在我的思维中勾勒出一幅注有非物理性精华物（比如

自我之叶或心灵活力）的物理实体的画面，我在无意中又倒退回从纯粹物理世界中衍生的意象。所以，对我而言，试图想象“一团意识”或者一个“非物理性的灵魂”，则我的思维将不可避免地遭遇一个半透明的、发光的烟晕旋涡，漂浮在它所居的物理实体之内，可能会从四周泄露一点儿。请注意，我完全清楚这是绝对错误的想象，因为这个现象根据定义就不是一个物理现象。但是正如我所说，我的想象是贫乏的，我需要借助于这种物理性的拐杖。

不论如何，要说在充满了一团团的意识的实体和那些缺乏意识的实体之间，有一道泾渭分明的界限，这个观念必然会引出各种各样的难解的谜题，比如下面这些：

> 哪些物理实体拥有意识，哪些没有？是整个人类的身体拥有意识，还是只有人类的*大脑*拥有意识？还是说只有特定*部分*的大脑拥有意识？一个有意识的物理实体的精确边界如何划定？一种物理结构要具有怎样的组织或化学特性，才有权利获得一团意识的入驻？
>
> 意识这种捉摸不定的万灵药对某些物理实体趋之若鹜，而对另一些却唯恐避之不及，那么从本质上决定这种倾向的是何种机制？意识到底具有什么样的奇妙模式识别算法，能令其万无一失地识别配得上它的物理实体，从而得以让自己投身其中？
>
> 意识如何得知要做这件事呢？它会设法环游物理世界，一路搜寻其可以投身其中的候选物理实体吗？或者它会向世界投下隐喻的闪光，一片一片地扫视检查，间或自言自语道：“啊哈！这里*有*一个实体，配得上标准剂量的一团我！”
>
> 意识是如何附着到特定的物理结构之上，而不会不小心沾到附近的物质呢？用来完成这种附着的是哪一种“胶水”？这种“胶水”有没有可能年久失效，让意识不小心脱落或者转移到了别的东西之上？
>
> *你的*意识跟*我的*意识有何不同？是不是我们各自都带有不同的

序列号或者“风格”，从而在我们之间搭建起一道密不透风的隔墙？如果你的那团意识附着在我的大脑上，而我的那团附着在你的大脑上，那么你会是写下这些话的人，而我变成了读者吗？

意识如何跟物理法则共存？也就是说，既然物理法则单凭一己之力便足以决定事物的行为，那么当一团意识在使唤物质材料的时候，如何才能避免跟这个事实发生激烈的冲突呢？

心灵活力的滑尺

读到这里，有些读者可能会说，我没有给予心灵活力（意识）足够的尊重。他们可能会说，这种精华物的分配是有渐变等级的：有些实体接收到很多，而另外一些实体则得到很少，或者干脆没有。它不是全或无的；相反，附着在任意给定物理结构上的意识量都并非精确的一个团，而是可以拥有任意多个团（不一定是整数个）。这无疑是一种改进！

可是话说回来，对于这一部分读者，我仍然为他们准备了不少问题，比如下面这些：

附着在一个给定物理实体上的意识有多少个团（或者百分之多少个团）是如何精准确定的？与此同时，这些团都储存在哪里？换言之，意识的“中央银行”在哪里？

当一定量的意识已经分配给一个受体（罗纳德·里根、一台下国际象棋的计算机、一只蟑螂、一个精子、一朵向日葵、一支温度计、一堆叶桩、一块石头、开罗城）后，这种配给方案就算是板上钉钉了，还是说配给量在随着涉及受体的物理事件的发生而不断变化？如果受体在某种形式上发生了改变，它的配给量（或者其中一部分）是要归还意识的中央银行，还是就继续四处飘荡下去，不再依附于任何物理的港湾？而如果它是在无牵无挂地四处飘荡，它还

会保留自己曾经附着过的那个受体的残痕吗？

患有阿尔茨海默病和其他痴呆症形式的病人——他们到临终前还是跟从前有一样的意识吗？到底是什么能让某种东西在很长的时间里始终是“同一个受体”？“罗尼·里根”“罗纳德·里根”“里根州长”“里根总统”和“前总统里根”这些在几十年间不断变化的模式仍然是一个“单一的实体”是由谁，或者由什么裁定的？而如果不管它变得多么易逝和易碎，它毫无异议地在客观上真的*就是*一个单一的实体，那么这个实体凭什么不能至今*仍然*存在呢？

胎儿（或者它们正在成形的大脑，甚至只包含两个神经元就可以算数）的意识又该怎么说呢？牛（或者它们的大脑）呢？金鱼（或者它们的大脑）呢？病毒呢？

我希望上面列出的种种难题已经清楚地表明，一种名为意识或心灵活力的精华物将招致数不胜数的问题，层出不穷，无穷无尽。二元论的信仰，通向的是一个巨大而阴暗的谜团陷阱，无可救药。

宇宙Z中语义学上的吹毛求疵

我最后想要处理的一件事跟大卫·查尔默斯那个著名的僵尸孪生兄弟有关。还记得吗？这个宇宙Z中的大卫在声称它喜爱冰激凌和紫色的花时，真心*相信*它所说的话，可事实上，它说的却是假话：它根本无法喜爱任何东西，因为它根本感觉不到任何东西——不比摩天轮的零件在啮合和转动时感觉到的东西更多。好吧，这里让我最困扰的地方在于，竟然有人不问青红皂白地愿意说，这位完全没有感觉的大卫会*相信*某些事情，它甚至还*真心*地相信它们。真心的信念难道不是感觉的一种变体吗？摩天轮中的零件会真心相信任何事吗？我希望你的回答是不会。抽水马桶里的浮球会真心相信任何事吗？我希望你的回答还是不变。

所以，假设我们从真心那一点上往后退一步，只是说，宇宙Z的大卫相信它口中说的那些有关它喜欢这个欣赏那个的假话。好吧，这样就又绕回来了，难道不能说，信念是感觉的一种吗？我不打算在此进行这番争辩，因为那不是我在乎的要点。我的要点是，跟我们这个复杂世界上的那么多区分一样，在确乎包含感觉的现象与不包含感觉的现象之间那个显在的区分绝不是非黑即白的。

如果我请你写下一份清单，列出那些逐渐从完满的情绪和感知滑向毫不动情和无知无觉的词汇，我想你可能不费吹灰之力就能完成。事实上，让我们现在就来快速一试。我的脑海里蹦出了这样动词，粗略地按照情绪性和感知性由高到低的顺序排列如下：悲痛、狂喜、折磨、欣赏、欲望、倾听、听见、品尝、感知、注意、考虑、推理、辩论、声称、相信、记忆、遗忘、知道、计算、表达、记录、反应、反弹、转弯、移动、停止。我不能说，我这份极短的动词清单，排列顺序是严格无误的；我只不过是把它们凑在一起，试图展示其背后是否含有感觉的在场无疑是存在一条波谱的，或者说，有一个渐变的灰度。然后，棘手的问题出现了：这些动词（以及形容词、副词、名词、代词等）中的哪一些是我们愿意使用在宇宙Z中大卫的僵尸孪生兄弟身上的？有没有一条精确划定的分割线，把不允许使用的词汇清晰隔开？谁来决定那条分割线呢？

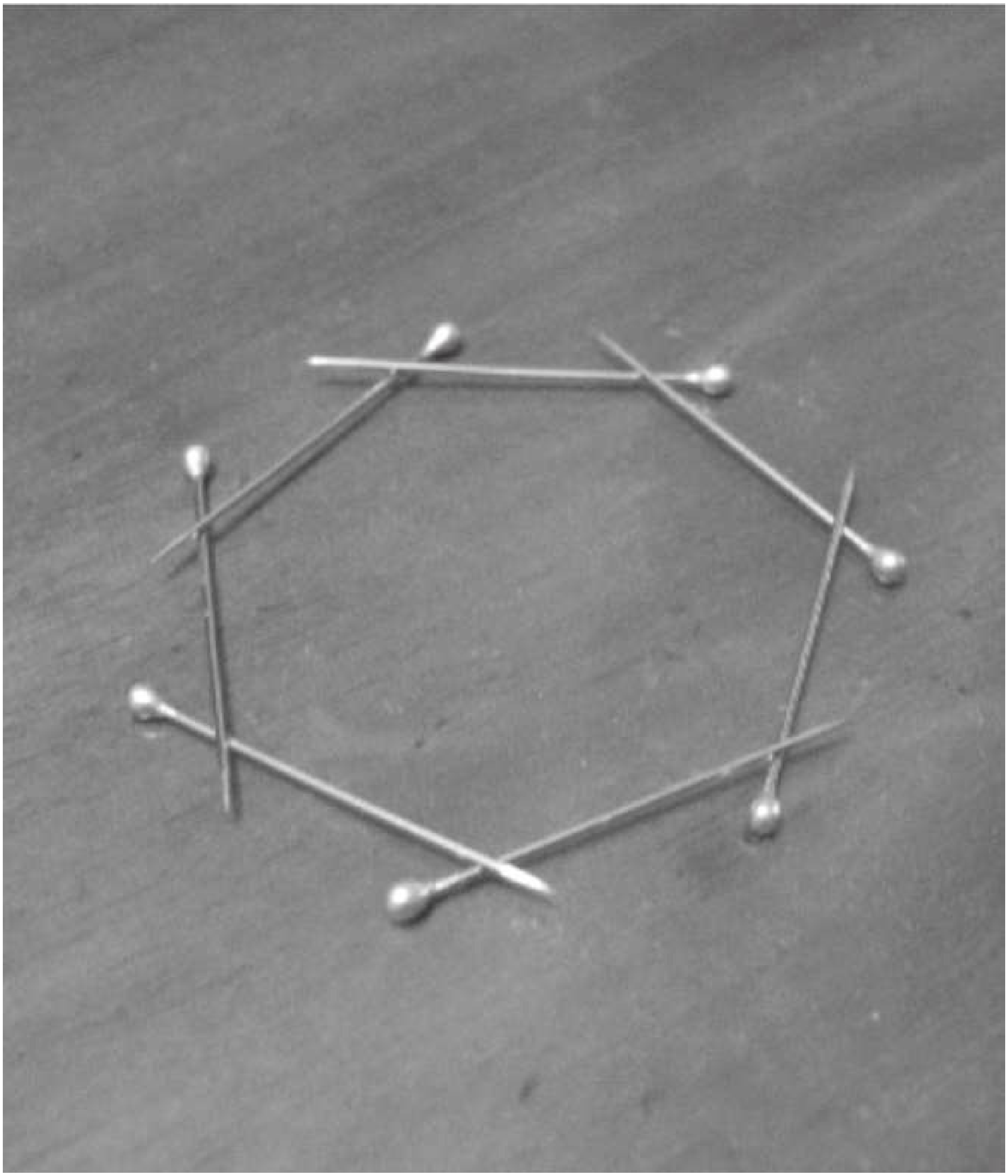
要想严肃地处理这个问题，不妨想想那个当我们在观看《星球大战》中类人机器人R2-D2和C-3PO的滑稽举动时在心中毫不费力（我最初写的是“无意识”，但是后来我觉得，在目前这个语境里，使用那个词显得太奇怪了！）就设定的标准。当他们中的某一个在我们认为是恰当的场景中表现出恐惧并试图逃离的时候，我们是否可以合理地使用“受惊的”这个形容词？或者，我们是否需要提前准备好对某种词语使用的限定，只有当构成目标行动背景的那个宇宙是一个充满了心灵活力的宇宙时，才允许使用这些词？一个宇宙的这种“科学”参数又该如何确定呢？

如果太空冒险电影的观众们在电影开始的时候就被“科学地”告知接下来的探险历程发生在一个跟我们的宇宙完全不同的宇宙里——也就是一个没有心灵活力的宇宙，那么当他们随后看到某个很像R2-D2或C-3PO（任你挑选）的可爱机器人被一个更大的机器人敲成碎片时，是否会完全无动于衷呢？父母会不会告诉他们抽泣的孩子们说：“安静！别哭了！那个愚蠢的机器人本来也不是活的！拍电影的人在一开始就告诉过我们，它活在那个宇宙里，而那里根本不存在有感觉的生物！一个都没有！”？活的（being alive）和活在（living）有什么不同吗？更重要的是，到底是什么引起了抽泣呢？

在宇宙Q中吹毛求疵

于是，在本章的结尾处，我们在绕了一大圈儿之后，又被带回了最开始的那个“学究式的语义学上的”代词问题。我们是否应该使用不同的代词来指代宇宙Q的大卫·查尔默斯（这个代词显然是“他”）和它在宇宙Z中彼此难分的那位僵尸孪生兄弟（这个代词很明显就是“它”）呢？当然，这种语义学上的吹毛求疵不仅限于人类和他们的僵尸孪生子。如果说在我们的宇宙里——我们那温暖、暧昧、洋溢着心灵活力的宇宙Q，毫无疑问，一只蚊子是一个可以拍死的“它”，那么火鸡呢？如果一只火鸡毫无疑问只是一道感恩节的晚餐菜肴，那么南美洲栗鼠呢？如果一只南美洲栗鼠只是一件皮外套，那么兔子、猫和狗呢？然后，人类的胎儿呢？新生儿呢？“who”（谁）和“which”（哪个）之间那条分割线到底在哪里？

正如我在本章开篇处所言，我认为这些都是很重要的问题——这些问题最终都是生死攸关的大事。这些问题可能不太容易回答，但是思考这些问题的答案是很重要的事。语义学并不总是学究式的吹毛求疵。



1. 令我觉得讽刺的是，大卫的很多关于意识问题的明确而细致的观点虽然和我的看法大相径庭，但却根源于我大概在15年前在印第安纳大学概念和认知研究中心的工作（至于641房间的那张橡木桌，那就说来话长了……）。大卫和我以及卡罗尔关系都很好，虽然我们关于感质、僵尸和意识的问题都存在分歧，我们依然是好朋友。

2. 此地为杜撰。——译者注

3. 我不打算就有关可怜的布鲁托的可能的身世（迪士尼的布鲁托是一只狗吗？）展开一场热切的辩论，但是我认为这个问题从认知科学的角度来看是非常吸引人的，因为它引出了一个关于人类思维中范畴与类比的本性的深刻问题。

4. See “Planet without Laughter” in [Smullyan 1980], a wonderful tale about vacuously laughing zombies.

5. See especially “The Unimagined Preposterousness of Zombies” in [Dennett 1998] and “The Zombic Hunch” in [Dennett 2005] for marvelous Dennettian arguments.

6. 事实上，这幅图像属于比尔·弗卢赫特，所以你可以用比尔的名字进行引用。我本来写了点关于飞侠哥顿的东西——时髦的斗篷扮相，而比尔可能觉得这幅20世纪50年代的图像是在是太过时了，甚至迂腐到可笑，所以及时地以一己之力把我拉回了21世纪。

7. I concocted the phrase “Capitalized Essences” when I wrote the dialogue “Three-Part Invention” in [Hofstadter 1979].

第23章 杀死两头神圣的奶牛

一条天蓝色的沙丁鱼

在关于意识的哲学文献中，所谓的“倒置波谱问题”常常让人头晕。在尽力精确地描述这头神圣的“奶牛”之后，我最好尽快屠宰掉它。（因为它患上了神圣的疯牛病。）

这一切的始作俑者是那个宣称你我内在之间绝无沟通的可能的主张——你完全无法知道我的内部是什么样的，反之亦然。具体而言，当你和我看到同一束玫瑰花时，我们都把各自所见外化成大概差不多一样的噪声（“红玫瑰”），但也许*我*在我那个私有的、封闭的闹壳内部体验到了红色，[注] 而如果你能够“进入”我的主观性内部驻足片刻，*你*很可能要说那是“蓝色”。（顺带一提，倒置波谱之谜的鼓吹者会拒斥任何一种暗示你和我真的已经*在*彼此内部的假设。谜题成立的前提是存在一个不可逾越的“你-我裂谷”。也就是说，一个人绝对无法进入他人的内在。换句话说，倒置波谱的信念是笛卡儿自我的近亲——认为我们都是互不相连的孤岛，而“你不能任意穿梭”。）

蓝，白，红=红，白，蓝

让我们再想想下面这个情况。也许，只是也许，当所有5 000万名法国人看到鲜血并声明它是“红色”的时候，他们其实体验到的是一种蓝

色的内在感觉；换句话说，在他们的眼里，血看起来就跟美国人眼中融化的蓝莓冰激凌一样。而当他们在夏日凝视一片美丽无云的天空并说出“蓝”这个字时，他们在视觉上体验到的其实是融化的红莓冰激凌。青天明鉴！他们受到了一场系统性的欺骗，与此同时，还有一轮系统性的语言掩护同步展开，阻止包括他们自己在内的任何人认识到这场骗局。

但凡我们能进入他们的大脑内部，用他们独特的“蓝-白-红”(注)体验颜色，我们就能确信这种倒置，可是，我们永远都做不到。他们也永远不可能用我们的“红-白-蓝”来观看颜色。更别说，在那些法国人的大脑里，并没有某种特殊缠结的线路——从任何尺度上看，他们的大脑都跟我们的大脑没有任何不同之处，无论是神经递质、神经元还是视觉皮层。它并不是某种可以通过神经重连或其他任何物理操作可以调整的东西。它只关乎某种不可言喻的感觉。更糟的是，虽然这件事是真的，但是永远没有人会知道它是真的，因为没有人能从一个人的内在溜进另一个人的内在——我们都被困在我们自己的头盖骨里。

这个设想听上去傻透了，不是吗？居住在一个相当随机的六边形国家边境线内部的5 000万人，刚好全部错误地把红色当成了蓝色，而又把蓝色当成了红色（可是从来没有在语言上透露出这一点，因为他们都学会了把那种蓝色的感觉称为“红”，而把红色的感觉称为“蓝”）。这怎么可能呢？

即便是最固执的倒置波谱拥护者，也会认为这个设想是荒诞无稽的。可是这同标准的倒置波谱其实一模一样；它只不过是把倒置波谱推广到整个文化的层面上，迫使它露出了本来的面目——像一则天真幼稚的童话。

倒置声谱

让我们调转另外的旋钮，再稍微进一步探讨一下倒置波谱。请设想，如果钢琴奏出的那些欢快的高音音符（亲爱的读者，我们都同意高音是欢快的，没错吧？）在比如戴安娜·克瑞儿（Diana Krall）的耳朵里听起来是非常低沉的（虽然她总是称它们为“高”音），而所有的低沉音符在她听起来都很欢快（虽然她总是把它们称为“低”音），那么又该怎么看呢？这还是一个倒置波谱问题，只不过涉及的是声音波谱，而不是视觉波谱。这个设想给我造成的印象，比原来那个关于颜色的倒置波谱更加不可理喻，而且我希望你也有同样的感觉。但是，在一个声音波谱和一个视觉波谱之间，为什么会存在根本性的差别呢？

首先，一个相当清楚的事实是，随着音符音高的下降，构成它们的每一次单独的振动都变得越来越清晰可辨。如果你按下钢琴最左边的琴键，当你（差不多）感受到它的音高时，你还将同时感觉到一阵急速的脉动。这个音符太低了，我们已经触及了一条临界线，一边是把它作为一个统一的音高来听，一边把它听成——或者不如说是感受成——一系列快速的振动。这个低沉的“音符”游荡在单数和复数之间的某处，徘徊在可听与可感之间。如果在我们的钢琴上，这个琴键左边还有额外的15或20个琴键（某些贝森朵夫（Bösendorfers）钢琴会多加一两个琴键，但我们想要的那架钢琴远比它们激进多了），那么这些超低的音符会开始让人感觉更像是我们肌肤和骨骼的振动，而不再是音高。当我们敲击两个相邻的琴键时，它们不会发出可以区分的音调，只会制造出低沉、粗壮的隆隆声，低深、绵长，感觉像是一阵闷雷，像是远处发生的爆炸，或者像是汽车掠过身边时鸣响的低音喇叭，爆发出令人惊奇的野性的震颤，而不是可以唱出来的音高。

一般而言，低音音符在不断降低的过程中会令人难以察觉地滑向身体的振动，而不再是音谱上的音高，而高音音符在不断升高之后，却没有这种变化。这就在音频波谱的两端之间建立起一种简明而显眼的客观差别。正因如此，说戴安娜·克瑞儿能够拥有一种倒置波谱的体验——也就是能够在低音琴键被敲响时体验到你我口中的高音——才令人难以

置信。毕竟，高音音符不可能造成客观的身体振动！

格叻冰和呐尔金

好吧，问题来了。如果有关一种声音的倒置波谱的观念是不合逻辑的，那么视觉的倒置波谱凭什么看上去更加合理呢？与可听范围内的声音波谱一样，视觉的电磁光谱波段的两端也是在物理上彼此不同的。一端是低频的光，可以被特定的色素吸收，而另一端则是高频的光，可供另外的色素吸收。但是，跟轰鸣声不一样，这些细胞运载的色素对我们而言只是存在于知识层面的抽象物，这给某些哲学家造成了一种错觉，以为我们对于红色和蓝色的体验完全与物理学没有关系。他们的结论是，一种颜色的感觉只是某种个人的发明而已，而两个不同的人可以做出不同的“发明”，谁也不比谁更有道理。

为了表达得更清楚一点，让我们假设，格叻冰和呐尔金（都是我刚刚生造出来的词）是每个人类大脑都可以享用的两种极其不同的感受。所有的人类都还在娘胎时就已把这些体验收入了他们内在的存储库里。你和我生来就有格叻冰和呐尔金，这是我们的标配特征，而自我们的婴儿时期以来，我们已经不下千万次地体验过这两种感受了。可是，对有些人来说，红光让他们感受到格叻冰而蓝光让他们感受到呐尔金，而对于另外一些人而言，情况正好相反。在你很小的时候，蓝色或红色中的一种刚好更多地触发了呐尔金，而另一种更多地触发了格叻冰。到五六岁左右，这种最初的倾向就永久定型了。科学无法预测过程和结果——但它就会这样发生。因此，亲爱的读者，你和我也许最终落在了格叻冰/呐尔金隔离墙的两边——但谁知道呢？谁可能知道呢？

我必须强调，在这个倒置波谱的情境中，红光（或蓝光）与呐尔金的联系不是某种在一个婴儿的大脑里启动，并随着大脑的成长而加固起来的后天的神经关联模式。事实上，虽然我在前面说，格叻冰和呐尔金

是所有的婴儿大脑里都天生具备的体验，但是它们并非是可以区分的大脑过程。不管用多么先进的大脑扫描设备都不能确定我的大脑（或者你的大脑）到底是在感受格叻冰还是呐尔金。简言之，我们谈论的并不是有关大脑的客观可见或可以测量的事实。

如果所有的倒置波谱之谜围绕的都是客观可见的事实，那么想要分辨出我们自己和那5 000万名内部感觉大错特错的法国人就太简单了！我们只需要检查他们的大脑灰质，然后精确定位那个说谎点就可以了；在那一点上，某些关键的联结跟我们大脑中的方向是相反的。接下来，我们还可以目睹他们的法国人大脑发生格叻冰的体验，而同样的刺激在我们的大脑内激活的是呐尔金。可是这些都与倒置波谱观念的真正意思背道而驰。它的意思是，尽管拥有完全一致的大脑神经联结，看见同一颜色的两个人还是能拥有完全不同的视觉体验。

倒置的政治波谱

这个假设性的概念让我们对彩虹光谱内颜色的内在体验听起来像是一套先在的纯粹抽象物，跟我们大脑外部乃至内部的物理学没有紧密的（事实上是根本没有）关系；恰恰相反，这些内在体验与外部的现象之间的映射是随机的。随着我们的成长，彩虹内的颜色映射到预先编织好的感觉光谱之上，这是我们大脑的“出厂”设置，可是这种映射并不是以神经联结为媒介实现的；说到底，神经联结是可以从一个疏离的第三人称视角观察到的，比如一位神经外科医生的视角，所以这种情况就被排除了。

下面，让我们仔细想一想如果主观感觉与外部刺激相互独立的话，那将意味着什么。也许可以随便举出一个例子，“自由”这个抽象概念给我的感觉跟“束缚”这个抽象概念给你的感觉一样——只不过，我们都在用“自由”这同一个词来表示它，因而我们误以为这对我们来说是同样的

体验。这听起来不太可能，是吗？毕竟，自由是令人愉快的，而束缚是令人不愉快的。但是话又说回来了，谁能说准呢？也许，令我感到愉快的体验对于你而言是不愉快的，反之亦然。

也有可能，当我撞入狂热的右翼摇旗者和反对堕胎者（他们在2004年美国总统大选中统治了“红”州）的阵营中时，我在内部感受到的那种翻腾的感觉，跟你在遭遇左翼焚旗者和支持堕胎者（他们在2004年的美国总统大选中统治了“蓝”州）时感受到的那种内心翻腾的感觉是一模一样的，而反之亦然。这难道不就成为一条倒置的政治波谱了嘛！你现在是不是有点儿迷糊了？（也许你体验到的迷糊，在我的体验中却是清晰分明。可是，让我们别再往下说了。）

那些把倒置的视觉波谱完全当真的哲学家，绝不会严肃地看待倒置的政治波谱。但是为什么不呢？原因可能是，他们不认为我们的大脑在“出厂时”就在内部预先配置了政治的“感觉”，那种感觉会随着我们的成长随机地依附于右翼或左翼阵营。可是他们*确实*认为，我们生来就内置有呐尔金和格叻冰（虽然他们没有使用我的词汇）。

我想再次提醒你，呐尔金不是大脑中一种可以识别的物理现象（格叻冰也不是）。呐尔金是当红色的光（或者，如果你是法国人的话，那就是蓝色的光）照在你的眼睛上时，你理应产生的那种内在的、不可言传的*感受*。法国人大脑里发生的内部物理事件，跟我们大脑里发生的一模一样，但是他们没有产生跟我们相同的*体验*。当红色的光刺激他们的视网膜时，法国人体验到的是格叻冰，而当蓝色的光照射时，他们体验到的是呐尔金。所以说，如果这种呐尔金体验不是大脑中任何可以在物理上辨识出来的东西，那么它*究竟*是什么呢？

倒置波谱主义者说它是纯粹的*感觉*。既然这种区分完全与物理学不相干，它也就等同于二元论了（实际上，这是我们已然了解的东西，因为对于笛卡儿自我的信仰就是一种二元论）。

紫罗兰是红色的，玫瑰是蓝色的

为什么，那些咬住倒置波谱不放的人总是咬住沿着一维数值刻度分布的体验不放呢？把自己局限在红蓝互换之间看起来太缺乏想象力了。如果你认为跟别人说“也许，你对于红色的私人内在体验跟我对于蓝色的私人内在体验是一样的”这句话是合乎逻辑的，那么为什么说“也许，你看一枝红色玫瑰的私人内在体验跟我看一枝蓝色紫罗兰的私人内在体验是一样的”就不合逻辑了呢？

在一条波谱内部打乱颜色的观念，到底有什么神圣不可侵犯之处？为什么不能把所有种类的体验随意打乱重排呢？也许你对于红色的私人内在体验跟我听见钢琴上超低音符时的体验是一样的。或者，也许你前往观看一场棒球比赛的私人内在体验，跟我前往观看一场橄榄球比赛的私人内在体验一样。可还是要说，也许你前往观看一场棒球比赛的私人内在体验，跟我坐过山车时的私人内在体验一样。或者，也许跟我打开圣诞礼物包装时的私人内在体验一样。

我希望，这些话在你听来都是荒诞无稽的，这样你就可以一步一步地往回倒退，从这些变体返回到最初的倒置波谱之谜的设定，始终保持住这份荒诞感。我最满意的情况也不过如此了，因为在最初的谜题和它刚刚演化出的那个荒诞至极的版本之间，我没有看出任何根本性的差别。

一条猩红色的沙丁鱼

倒置波谱之谜的提出依赖于这样一种观念，即我们都生来具有一套特定的“纯粹的体验”，它们没有物理基础，却随着我们的成长附着在特定的外部刺激之上，于是特定的体验便与特定的刺激联姻了，从此以后一辈子都亲密地绑定在一起。但是这些“纯粹的体验”并非大脑的物理状

态。倒不如说，它们是一个人单纯“拥有”的主观感觉，不需要存在任何物理学上的解释。你的大脑状态和我的大脑状态可以看起来跟任何人想象中的都一模一样（即便使用高分辨率的大脑扫描仪也看不出差别），但是当我在感觉到蓝色时，你就是在感觉到红色。

倒置波谱的童话是一个鲁莽与胆怯的孱弱混合体。它一方面虚张声势地否定了物理世界与我们内部感觉的相关性，另一方面又唯唯诺诺地把自己圈定在一维波谱的藩篱之中，而且还仅限于那个电磁光谱。声音波谱跟震颤和振动等客观物理事件的关联太过紧密，让我们很难想象它如何能倒置过来，而如果一个人想把这个观念推及一维波谱之外的领域，它就会变得太过荒诞不经，根本无法令人信服。

是的，人们想要

在关于意识的哲学文献中，还有另外一种让我浑身起鸡皮疙瘩的东西，那就是所谓的“自由意志问题”[注]。让我来描述一下这第二头“奶牛”，然后也努力赶跑它，能多快就多快。（因为它也患上了神圣疯牛病。）

当人们决定去做某件事的时候，他们常常会说：“我是出于我自己的自由意志来做这件事的。”我认为他们这么说的意思通常在实质上相当于：“我做这件事，是因为我想要做，而不是因为其他人强迫我去做。”虽然“我是出于我自己的自由意志来做这件事的”这个说法让我感到很不舒服，但是我做出的那个同义转换解释，在我看来就非常自然。我们的确有想要的东西，而我们想要的东西也确实在导致我们去做事情（至少跟641的质数性导致多米诺链中的一张多米诺骨牌倒下的程度一样）。

生活的树篱迷宫

有时候，我们的欲望会撞上阻碍。别的人把冰箱里那最后一杯软饮喝掉了；以前通宵开放的杂货店现在到午夜就关门了；我朋友的汽车轮胎瘪了；狗把我的家庭作业啃掉了；登机口的舱门在30秒之前刚刚关闭；航班因为萨斯卡通的一场暴风雪而被取消了；我们的电脑出问题了，好像用不了PowerPoint了；我把我的钱包忘在了另一条裤子里；你看错了截止日期；面试官是厌恶我们的某个人；她听说这次工作机会的时候，已经太晚了；隔壁跑道的运动员比我跑得快。如此等等。

在这些情况下，我们的意志虽然在推动我们，可是已经不足以让我们得到我们想要的了。它把我们推向某一个特定的方向，但是我们却是在一座树篱迷宫的内部移动，哪条路走得通，是由这个世界的其他部分决定的，而跟我们想要什么无关。所以说，我们是在迷宫里不情不愿地移动，而不是自由自在。来自外部和内部的压力组合在一起，共同判定了我们在这座名为“生活”的疯狂树篱迷宫中的路径。

关于这一点，没有什么好迷惑的。而且我要重申，这些压力中有一些是我们想要的，这个观念也没有什么让人困惑的地方。说不通的是，除此之外，还有人坚持认为，我们想要的不知怎么就是“自由的”了，或者我们的决定莫名其妙地成为“自由的”。它们可是我们脑子里物理世界的产物啊！怎么可能是自由的呢？

根本就没有自由意志这么一回事儿

当一条公狗嗅到了一条发情的母狗的气味时，它会产生某种极其强烈的欲望，而且它会极尽所能地想要满足这个欲望。我们对于这种欲望的强烈性看得太清楚了，当这个欲望遭到了阻挠（比如，被一堵篱笆隔开或者被狗链拴住），我们都会心有不忍，同情那可怜的动物，它受困

于内在的驱动，被它根本无法理解的一股抽象力量推动着。这幅令人心酸的景象清晰地展示了何为意志，可那是自由意志吗？

我们人类真的拥有任何超越了狗的那种渴望的东西吗？我们也有强烈的渴望——其中有一些在性的竞技场中上演，还有一些活跃在更崇高的生活领域之内。而当我们的渴望得到满足时，我们会达到某种幸福的状态，但是当它们受阻时，我们就会跟那条被拴住的狗一样，处境悲凉。

那么，所有关于“自由意志”的那些大惊小怪到底有什么意义？为什么那么多人都在坚守着那个浮夸的形容词，甚至常常在那之中发现了人性的至高荣耀？如果“自由”这个词是准确的，它又会给我们带来什么？或者不如说，它有可能给我们带来什么？我真心不知道。在这个复杂的世界里，我看不到任何给我的意志“自由”的空间。

我很高兴自己拥有意志，至少当它没有被困住我的树篱迷宫严重挫败的时候，我很高兴能拥有意志，但是我不知道如果我的意志是自由的，那种感觉是什么样的。那到底意味着什么呢？意味着，我有时候没有遵从自己的意志吗？好吧，可是我为什么会那么做呢？为了折磨自己吗？我想，如果我想折磨我自己，我可能会做出这样的一个选择——可是那样一来，这就是因为我想要折磨自己，是因为我在元层面上的欲望比我表面上的欲望更强烈。所以我可能选择不多加一份面条，而不顾我——更应该说是一部分的我——想再吃一点儿的欲望，因为另一部分的我想要我不要变胖，而这个监视体重的部分恰好（在今晚）拿到了比那个贪吃的部分更多的选票。如果它没有拿到更多的选票，那么它就会输掉，而我内在的食欲则会胜出，我也将遵从这个支配了大脑的欲望。

没错，确定无疑的是，我会做出一个决策，而且我会通过进行这样一种内部投票来做出决策。票数会给出一个结果，而其中一方会切实胜出。但是，这里哪有“自由性”呢？

说到这里，那个与我们的选举过程进行的类比已经太过明目张胆㊟、欲盖弥彰，我应该把它说清楚了。这并不是说，在一个大脑里，仿佛有某种“神经选举权”（“一个神经元，一张选票”）；而是说，在这个大脑的更高组织层级上，存在某种“欲望层面的选举权”。既然根据我们对于大脑的理解，我没有办法把这些选举权精准地进行物理定位，我只能说它实质上是“一种欲望，n张选票”，而其中的n是与某种给定欲望联系在一起的权重。并非所有的n值都是等同的，也就是说，并非所有的欲望都生来平等；大脑不是一个平等主义的社会！

总的来说，我们通过一种民主社会投票的类比过程而做出决策。我们各种各样的欲望喧哗不止，还要考虑到很多外部因素的制约，或者用更具隐喻色彩的方式说，那些外部因素扮演着庞大的生活迷宫中树篱的角色，而我们正受困于此。生活的很大部分都是随机的，甚至到了令人难以置信的程度，而我们对此完全没有控制力。我们可以随心所欲地去发挥意志，可是大部分时间里，我们的意志都在挫败之中。

我们的意志不仅不是自由的，而且恰恰相反，它是稳定而恒常的，像一个内在的陀螺仪，而正是我们不自由的意志的稳定性和恒常性，让我是我，让你是你，并让我保持是我，而让你保持是你。童话里的自由人鱼只不过是另一头蓝色座头鲸罢了。



1. The most penetrating discussion of the inverted-spectrum riddle that I have read is that in [Dennett 1991].
2. 法国国旗的颜色依次是红色、白色和蓝色。但是法国人习惯倒着说成“蓝色、白色和红色”。这种无意的事例似乎让我们觉得法国人对于颜色的感知与我们“恰好相反”。
3. 我和丹尼尔·丹尼特总归要在某些地方出现观念不一致的地方，而在我这本书接近尾声的地方，我们终于找到了这个点。那就是关于自由意志的问题。我对于丹内特在1984年那本书中的绝大多数观点都持认同的态度，但是我难以赞同他认为我们拥有自由意志的说法，不管是何种自由意志。总有一天，丹和我会在两个人之间决一死战。
4. This idea of “votes” in the brain is discussed in Chapter 33 of [Hofstadter 1985], as well as in the Careenium dialogue, which is Chapter 25 of the same book.

第24章
论宽宏与友谊

灵魂有大有小吗？

我曾在本书的第一章中引用过詹姆斯·亨内克有关“小灵魂的人”的古怪警告，此后时不时地会在书中间接提及，漫不经心地指出构成不同人类灵魂的“亨内克”数，但是我还从来没有明确地说过，一个高亨内克或者低亨内克的灵魂更倾向于展现出何种特征。确实，任何有关这种区分的暗示，都有激发众怒的风险，因为在我们的文化里有一种教条，粗略地说，就是所有的人类生命都有绝对等量的价值。

尽管如此，我们还是每天都在习以为常地违反这条教义。最明显的例子，就是在一场公开宣战的战争中，我们作为一个社会群体公开地滑入另一种集体模式之中：一个庞大的人类子集的生命价值在此突然之间缩减为零。我不需要对此详加说明了，因为它实在太过残酷了。另外一则明显违背我们教条的例子是死刑，社会群体集体选择终结一个人的生命。从根本上说，这就是社会集体判定了某个特定的灵魂不再值得任何尊重。即便没有死刑，还有监禁，社会以不同程度、或有或无的尊严对待不同的人，隐含地表现出对不同尺寸灵魂的不同等级的尊重。再想想外科医生在尝试拯救生命时采取措施的显著差异。突发心脏病的国家领导人（或者任何一家大企业的老板）接受的治疗，要远比一个平头老百姓好得多，更不用说非法移民了。

为什么我会把这些社会的不公平对待方式视为灵魂价值之间心照不宣的差别呢？因为我觉得，不管我们是存心还是无意，我们都在把一个

生命的灵魂尺寸等同于那个生命的“客观”价值，也就是我们这些外在的人给予那个生命内在的尊重程度。而我们当然不会给所有的生命赋予同样的价值了！我们会毫不犹豫地在人类和动物的生命之间，以及不同“等级”的动物的生命之间做出明显的区分。

所以，绝大多数的人类都在直接或间接地甘心参与杀害多种动物并食其骨肉的行为当中（有时候，甚至还在一个盘子里混合了猪、牛、羊的肢体碎片）。我们还冷酷无情地把我们杀掉的动物的身体碎片喂给我们的宠物吃。显然，这些行为在我们的思维里建立了一套动物灵魂国度中的等级秩序（除非有人用食古不化的非黑即白论调声辩说，“灵魂”这个词本来就不适用于动物，但是这种绝对主义在我看来更像是一种伸手拿来的教条，而没有经过严肃的思考）。

我认识的绝大多数人都会（要么通过言语明确表达，要么通过行为间接表示）把猫的灵魂排在比牛的灵魂更高的位置上，牛的灵魂比老鼠的灵魂更高，老鼠的灵魂比蜗牛的灵魂更高，蜗牛的灵魂比跳蚤的灵魂更高，以此类推下去。所以我问自己，如果不同物种之间灵魂尺寸的区分是如此寻常而没有威胁的一个概念，那么我们为什么不愿意考虑在一个单一物种之内，特别是我们自己的物种之内，拉出一条明确的（而不仅是默认的）灵魂尺寸的谱系呢？

从深度到高度

我在前面的章节中把自己逼入了一个死角，现在要想办法走出窘境，向这种区分报以粗暴的反击。我在完成这项任务时，将只取一条宽广波谱的两端，而亲爱的读者，你和我假定位于波谱中段的某个地方（但是，希望我们更接近“高”的一端，而不是“低”的一端）。

然后，在最低的一端，我安置的是无法控制自己的暴力精神变态者

——那些实质上没有能力内化他人（或动物）的心理状态的成年人，他们因为缺乏这种能力而惯常性地对其他生命犯下暴力罪行。也许这些人仅仅是不幸生来如此，但不论原因如何，我都把他们划分到波谱的最低端。直白地说，这些人的意识不如正常的成年人，也就是说，他们拥有更小的灵魂。

我不会给出一个量化的亨内克值，因为那将把我们置于荒诞可笑的境地。我只是希望你能理解我大概的观点，而且不会觉得它是不道德的。说到底，这个观点与主张这些人应该被关在监狱里的说法没什么不同，在我认识的人里，没有人认为监狱是本质上不道德的机构（当然，它们运转的方式就是另外一回事儿了）。

那么，波谱的更高一端呢？不出意外的话，我想我会指向那些与暴力的精神变态者在行为实质上正好相反的人。这指的就是像莫罕达斯·甘地（Mohandas Gandhi）、埃莉诺·罗斯福（Eleanor Roosevelt）、罗尔·瓦伦堡（Raoul Wallenberg）、让·穆兰（Jean Moulin）、特蕾莎修女、马丁·路德·金和塞萨尔·查韦斯（César Chávez）这样的人——这些卓越非凡的个人[注]在他们对于受难者的同情的指引下，把大部分生命都投入到帮助他人的事业中，而且是以非暴力的方式来行事。我想，这些人就是比正常的成年人更有意识的人，也就是说，他们拥有更大的灵魂。

虽然我极少重视单词的词源，但是当我在几年前为有关这些观念的一场讲座做准备的时候，我很欣然地注意到“宽宏”（magnanimity）这个词虽然对于我们而言实质上是“大度”的同义词，但是在拉丁语里，最初的意思是“拥有一个伟大的灵魂”（拉丁语中“animus”的意思是“灵魂”）。多亏了这种X光式的透视，我得以用一种新的眼光重新审视这个熟悉的词语，这给我带来了很多乐趣。（当我在整理这本书相当疯狂的索引时，意外地发现“圣雄”（Mahatma）——通常给予甘地的尊称——也有“伟大灵魂”的意思。）另外，“同情”（compassion）的词源也很有意思，它来自拉丁语，词根的意思是“随之经历苦难”。这些隐藏的信息

穿越千年的回响刺激我展开更进一步的探索。

阿尔贝特·施韦泽的宽宏

在我个人的心目中，伟大灵魂的典范是神学家、音乐家、作家和人道主义者阿尔贝特·施韦泽（Albert Schweitzer）。他于1875年出生在阿尔萨斯省的小乡村凯泽尔贝格[注]。1913年，他在加蓬（Gabon）的兰巴雷内（Lambaréné）创办了一所医院，在那里工作了50多年，并以此闻名世界。

早在很年轻的时候，施韦泽就懂得认同他人，给予他们怜悯与同情，并想要分担他们的苦难。这份充满移情力的慷慨之心从何而来？谁知道呢？比如，上学的第一天，六岁的阿尔贝特就注意到，父母用精致的服装把他打扮得比同学们都好，这种差异给他带来了极大的困扰。从那天起，他坚持打扮得跟比自己穷的同学一样了。

施韦泽的自传作品《我的童年与青年时代》（*Aus meiner Kindheit und Jugendzeit*）描绘了这种弥漫在他整个生命中的同情，下面是从中摘取的一段活灵活现的篇章：

> 就我所能窥见的童年而言，我经历了身边所见过的一切悲苦。我真的从来都不知道什么是简单而青春洋溢的生活之乐（joie de vivre），而且我相信，很多孩子都跟我一样，尽管他们在外面表现得仿佛拥有十足的无忧与快乐。
>
> 特别是，可怜的动物们不得不经受那么剧烈的痛苦与匮乏，这个事实让我备受煎熬。我看到一匹年老而跛脚的马被一个人拖着向前走，而另一个人则用手里的棍子打它，送它赶往科尔马（Colmar）的屠宰场，这一幕阴影在我的脑海里萦绕了几周的时间，挥之不去。甚至在我步入学校之前，我就已经发现，我在晚祷

> 时只应为人类的福祉祷告这件事本身就是说不通的。所以我会悄悄地念出自己编造的一段祷告词："亲爱的上帝啊，请保佑并祝福一切呼吸的生命，令其远离邪恶，温柔入眠。"

施韦泽对于动物的同情并不仅限于哺乳动物，而是一直向下延伸到波谱的低端，直至蚂蚁和蠕虫。（我说"一直向下"和"低"，并不带有轻蔑的意思，只是为了指出，施韦泽跟几乎所有人类一样，一定也持有一种"意识倒金字塔"的观念，跟我在前文展示的差不多。这样一种思维等级，既容易引发轻蔑感，也容易引发关心与责任感。）他曾对一个即将踩死一只蚂蚁的10岁男孩说："那是属于我个人的蚂蚁。你要是弄坏它的腿，是要赔钱的！"他会习惯性地把他在马路中央看到的蠕虫或者在一个池塘里挣扎的昆虫捡起来，然后把它放归田野或一株植物上，好让它有机会继续生存下去。实际上，他对于自己这种做法的解释是相当尖刻的："每当我帮助一只昆虫摆脱困境时，我都是在试图缓解人类在对动物犯下的罪行中染上的自责。"

众所周知，"敬畏生命"是施韦泽提出的一个简单而又深刻的指导性原则。当他在1953年获得诺贝尔和平奖时，他在获奖演说中宣称：

> 人类精神未死。它还在偷偷地活着……我们相信，所有的伦理都必须以同情为根本，而只有当同情对一切生物敞开怀抱，而不仅仅局限于人类之时，它才能在宽度与深度上达成圆满。

下面这则逸事同样取自《我的童年与青年时代》，而格外发人深省。春天，当复活节来临之际，七八岁的小阿尔贝特接到一个战友的邀请——毫不夸张地说，是荷枪实弹的战友，去参加一场猎杀鸟类的冒险行动，武器就是他们已经共同制作好的弹弓。几十年后，施韦泽回头再看这个生命中的转折点时，这样回忆：

这是一个令人厌恶的提议，但是我不敢拒绝他，因为我怕他嘲笑我。很快，我们就来到一棵光秃秃的大树附近，树上的小鸟在晨光中叽叽喳喳地欢快地叫着，对我们毫无戒备。我的同伴深蹲俯身，像一个印第安猎手一样，把一块石子装在了弹弓的皮袋里，拉紧了弓子。屈从于他投向我的专横目光，我也做出了同样的动作，内心与一阵阵良心的涌动斗争，与此同时对自己暗自发誓，他射击的时候我自己绝不会跟着他一起做。

就在那时，教堂的钟声忽然奏起，与阳光中鸟儿的鸣唱交响辉映。那是提前于主钟声敲响半个小时的时候奏响的早钟。但是，在我的耳中，这就是来自天堂的声音。我把我的弹弓丢掉了，惊吓到了那些鸟儿，于是它们飞到了一个安全的地方，远离我同伴的射程。而我也逃回了家中。

自那天起，每当圣周的钟声响起并回荡在春天叶芽未出的树林间，我都会怀着深深的感激，想起在那改变命运的一天，那钟声是如何在我的心中奏响了那句诫令：“不可杀生。”从那天起，我发誓要把自己从对他人的畏惧中解脱出来。每当我内心的信念受到挑战时，我对他人意见的看重都比之前减轻了很多。而且，我也竭尽所能地克服了自己害怕被同伴嘲笑的恐惧。

我们在这里看到了同伴压力与内在声音之间的典型冲突，或者用我们通常的说法（也是施韦泽自己的说法），就是一个人的良心。幸运的是，在上面这则例子中，良心以明显的优势胜出了。而且确如他所说，这个决定持续了一生之久。

是良心构成了意识吗？

在语义空间的这个领域内，还存在另外一个语言学现象让我感受到了最强的刺激。那个事实就是，在罗马语系中，“良心”（conscience）

和“意识”（consciousness）这两个在说英语的人看来截然不同的概念，本来却是用一个同样的词来表示的。例如，法语单词“conscience”就同时包含了这两种意思，而这是我在10多岁的时候买到的一本题为《大脑与意识》（*Le cerveau et la conscience*）的书[注]中第一次得知的。这也许仅仅是那些语言中一个词汇学上的断裂或者混淆的语义模糊现象（这个单词在字面上的意思是“共同的知识”），可是即便如此，我还是认为它为我们提供了一种在别处看不到的独特洞见：能够部分地内化其他生物的内在（良心）是把拥有大灵魂（很多意识）的生物与具有小灵魂的生物，以及没有或几乎没有灵魂的其他生物区分开来的最鲜明而可靠的标记。

蚊子没有良心，同样也没有意识，因而也就不具有任何配得上“灵魂”这个词的东西；我认为这是显而易见的，或者是几乎没有疑问的。这些飞行的、嗡嗡叫的、吸血的自动飞行物，更像是微型的跟踪导弹，而不像是有灵魂的生命。你能想象一只蚊子体验到慈悲、怜悯或友情吗？说到这儿就够了。下一个！

比如说，狮子——典型的食肉动物——情况又如何呢？狮子会追踪、猛扑、撕裂并吞食仍在踢腿、嘶鸣的长颈鹿和斑马，而它们这么做的时候，并没有一丁点儿的慈悲或怜悯之心，这说明它们完全缺乏同情。可是它们似乎对自己的幼崽却投入了巨大的关怀，抚摸它们、养育它们、保护它们、教导它们。这跟蚊子的行为太不一样了！除此之外，我怀疑狮子们可以很轻易地给予某个其他物种（比如人类）的幼崽关怀和照顾。从这种意义上说，一头狮子尽管对大多数其他生物的内在完全无视并漠不关心（令人郁闷的是，这种特质听起来像极了大多数的人类），它们还是可以而且愿意内化至少某些其他生物（特别是其他的狮子，尤其是它的直系亲属）内在的特定而有限的方面。

大多数的狗都关心其他生物——特别是属于它们圈子内部的人类。我想这也是显而易见，或者几乎毫无疑问。确实，众所周知，有些狗表

现出了令人难以置信的宽宏气度，愿意为它们的主人牺牲自己的生命。我还没有听说过有哪头狮子曾经为其他其他物种的动物做过类似的牺牲，但是我猜，也许曾经在某处，也有某种跟狗类似的狮子，为了拯救一个人类同伴的生命，而跟另外一头野兽搏斗至死。但是，要我们想象一头狮子选择去做一个素食主义者，就有点扯得太远了。

话说回来，只要去互联网上快速搜索一下就能发现，素食狮子这个概念一点都不罕见（需要承认的是，这通常都出现在虚构文学中，但也不总是这样）。事实上，有这样一头名为“小泰克”（Little Tyke）的母狮子，在西雅图附近被作为宠物养大。在长达四年的时间里（网上是这么说的），小泰克都拒绝吃任何摆在它面前的肉食，直到它的主人最后放弃了尝试，接受了它的素食主义立场，以及它在跟绵羊、小鸡和其他野兽嬉戏玩耍中表现出的快乐。直到生命结束，小泰克一直都是一头素食的狮子。奇迹从未停止发生！

不管怎么说，拥有良心——一种对于道德的认识以及一种对其他有知生命做“正确的事”的在意。在我看来，是标示着一个生命中存在意识的最自然的标志，也很有希望是最可靠的标志。也许，这可以简单地浓缩成一个生命在“金规则”的践行中投入的多少。

阿尔贝特·施韦泽和约翰·塞巴斯蒂安·巴赫

我不得不承认，我总是直觉地感受到，在意识的测量上，还有另外一种十分模糊又充满争议的标尺：音乐品位。我当然无法解释或者为我自己的音乐品位辩护，而且我知道，一旦我试图这么做，我就会陷入水深火热的旋涡之中，所以我根本就不会尝试。尽管如此，为了讨论阿尔贝特·施韦泽和他的音乐的深刻性，我还是不得不透露一点儿我自己的音乐品位。

在我16岁生日那天，我的母亲送给我一张音乐唱片，收录了巴赫具有丰碑意义的作品《平均律钢琴曲集》（*The Well-Tempered Clavier*）第一卷中的前八首前奏曲和赋格[注]，钢琴演奏者是格伦·古尔德（Glenn Gould）。这还是我第一次接触到“赋格”这个概念，而它在我年轻的心灵中产生了一种电击的效果。在接下来的几年里，我每走进一家唱片店，都会四处搜寻钢琴演奏版的《平均律钢琴曲集》的其他部分，因为它在那个年代真的非常罕见（即使是大键琴演奏版本也很罕见，而我更喜欢的钢琴版本更加罕见）。每当我又找到了一套新的前奏曲和赋格时，不管是哪一卷，我会把唱针放在新唱片的槽纹里。这个动作，以及我第一次听到那音乐时的感受，都可以列入我生命中最激动的事件。

在我父母的音乐收藏中，还有几首阿尔贝特·施韦泽演奏的巴赫风琴作品，但是我过了很久才愿意试着听一听，因为我担心它对于我而言太“重”了。但是当我最终听到的时候，我受到的感动无与伦比，对于它的痴迷绝不亚于我对于《平均律钢琴曲集》的痴迷。自此以后，我自然把在唱片店的搜索的范围扩大到了巴赫的风琴作品。可是我很快发现，自己受到了某件事的困扰，那就是，很多演奏者对风琴作品的处理都太过轻松和活泼了，仿佛它们仅仅是名家名手的练习曲，而不是关于人的境遇的深刻陈述。施韦泽的演奏是谦逊而简单的——让我着迷之处在于，他时不时地会犯错误，但还是不受干扰地继续演奏下去（在其他的录制音乐中，不会在任何地方出现哪怕一处错误，这让我感到非常不自然，甚至很古怪）。另外还有一点巧合的是，虽然我当时并不知道，可是那些演奏者全都是在阿尔萨斯省的根斯巴哈（Günsbach）的同一家教堂里，用同一架风琴录制的。那里的钟声曾在一个明亮的春季早晨奏响，拯救了一两只鸟儿的生命，并改变了年轻的阿尔贝特的人生，而后数千人的人生也因此而被改变了。

欣赏那深刻性！

随着时间推移，施韦泽演奏的巴赫已经深深地成为我的一部分。我又寻得了另外几张他的演奏作品，全部同属于一个系列，每一张都揭示出一种宇宙智慧的新深度（也许这听起来太夸大了，可是于我而言，这么说恰到好处），这种深度既来自作曲者，也来自演奏者。

《集异璧》这本书的流行，让我的名字在音乐群体内以某种方式和巴赫的名字联系在一起（这真的是极大的荣耀），我对此自然充满感激。而在1985年，巴赫的300周年诞辰之际，我还有幸参与了几场300周年纪念庆典，包括我在他生日当天组织的一场小小的活动。我在安娜堡组织的这场活动，面向的是我班上的学生和几位朋友。这场活动最为精彩的一瞬，是我们在我预订的巨大生日蛋糕上点亮了300根蜡烛时释放出的小型火焰风暴。

15年之后（2000年），我又意外接到邀请，去参加一场在意大利罗韦雷托（Rovereto）举办的巴赫250周年忌辰（巴赫于1750年7月逝世）的纪念活动。因为我正好要在那个时候前往意大利北部，所以我欣然接受了这份邀请。活动的当天下午有几场纪念讲话，在晚宴过后，又安排了一场表演——由一个知名的歌唱团演唱几首巴赫的作品（适合小型合唱的改编版本）。我之前对他们的歌唱技巧是有印象的，因而很期待当晚能欣赏一场感人肺腑的音乐盛宴。

可是，我听到的却是与期待完全不同的东西，虽然我其实早就该预料到这种情况了：无休无止的炫技，除此无他。那绝对是令人印象深刻的表演，可是在我的心目中，它却又是极其乏味的。在我看来，当那些歌手唱到巴赫所有风琴赋格中最深刻的一首——通常被简称为《伟大》（*The Great, BWV 532*）的G小调赋格——时，整场演出跌入了最低点。而阿尔贝特·施韦泽演奏的这首曲子，我却非常喜爱。他的演奏虽然低调而谦虚，但其中包含的感受在深度上是无与伦比的。令人懊恼的是，我再也没有忘记那些歌手是如何以大约两倍的速度演绎了这首富于冥想性的赋格：他们像赶火车一样地冲刺，然后目中无人地炫技扬威。他们

踮着脚尖摆动，仿佛在邀请观众跟着他们时髦的节奏一起摇摆。他们甚至还用手指打起了节拍（单是“节拍”这个词在这种神圣的语境里听起来就足够荒诞了）。有几位歌手偶尔会咧嘴向观众抛去明媚的一笑，好像在说，“我们是不是太棒了？你这辈子有听过任何人能在一秒钟唱出这么多音符吗？更别提那些颤音了！这音乐难道不性感吗？希望你们都懂得欣赏！而且别忘了，我们还有很多CD唱片，等表演结束之后，你们可以买回家哦！”

所有这些都让我感到格外的震惊。当然，在这个世界上，表演任何一种音乐作品的方式都有很多种可能性。在这些歌手的速度和炫技之中当然也有某些有趣的东西可说，而且他们把那种超速的颤音表现得毫无瑕疵——给人带来深刻印象的方式跟一部炫酷赛车的工艺细节一样，但是对我而言，它跟这首音乐的意义毫不相关。这首音乐的意义是沉思性的、宇宙性的，而不是任意的、显摆的。我对于演奏音乐的多样化形式是很包容的，但是我也有我的底线，而他们的表演远远越界了。它让我无比渴望听到阿尔贝特·施韦泽用根斯巴哈那小小的乡村风琴演奏的带有轻微瑕疵的、充满沧桑感和反思性的深刻性。但是在那天晚上，根本就没有这个可能。那是神圣与亵渎碰撞的典型，至今仍在我的脑海中恍如昨日。

我在准备这一章的写作时，才首次遇见施韦泽本人所写的一些文字；它们是我在罗韦雷托那晚遭受的巨大困扰的奇怪回声（要是回声能早于发声！）。下面就是他在几乎100多年以前写下的话，讨论的是那个时代对巴赫作品的演奏：

> 很多演奏者演奏了很多年的巴赫作品，却没有亲自体验过巴赫能够在任何一位真正的艺术家身上引起的那种深化。我们大多数的歌手都太过沉迷于技巧，以至于难以正确地演唱巴赫。只有很少的几个人能够重现其音乐的精神；其余的人都没有能力穿透到这位大师的精神世界之中。他们感觉不到巴赫想要说什么，因而也无法将

> 其传达给其他任何人。最糟糕的是，他们都认为自己是杰出的巴赫的诠释者，并没有意识到他们缺乏的东西。有时候，人们不禁会纳闷，表演如此肤浅，观众怎么可能从中感受到一丁点儿巴赫音乐的深度！
>
> 凡是理解当下境况的人都不会认为这些评论是过分悲观的。我们对巴赫的神迷正在遭遇危机。危险在于，我们对于巴赫音乐的爱将变得流于表面，让太多的虚荣与自负混在了里面。我们这个时代令人叹惋的偏向模仿的趋势，也蔓延到了我们对于巴赫的继承上——如今，这已经再明显不过了。我们表现得仿佛我们想要赞美巴赫，但是事实上我们只是在赞美我们自己。我们表现得仿佛我们重新发现了他，理解了他，并且以前所未有的方式演奏了他的作品。少一点噪声，少一点“巴赫教条主义”，多一点能力，多一点谦逊，多一点宁静，多一点投入……只有这样，巴赫才能获得比之前更多的在精神上和真理上的尊重。

我对于这段猛烈地批判肤浅性的文字几乎是全盘认同；我只想简单补充一句，读到这段话后，我得到了深深的宽慰，虽然我是在罗韦雷托事件的几年之后才读到的，但它仍然让我认识到我的哀叹之情并不孤独。施韦泽是那种最为谦卑和恭让之人，所以他的评论一定是对一个值得叹惋的趋势的最诚实的反应。那个趋势在一个世纪以前已经露出水面，而在今天看来，这个趋势只增未减。

凡是蚂蚱，必有一死

有些读者可能不禁发问，这一切跟“我”、意识或者灵魂有什么关系呢？我的回答是：“与让自己跟阿尔贝特·施韦泽和巴赫组合的精神性融合在一起相比，还有什么东西与意识或灵魂的关系<u>更大</u>呢？”

一天晚上，为了重温我有关施韦泽演奏巴赫风琴音乐的尘封记忆

（我在10多岁到20多岁之间曾听过上百遍），我从我的架子上抽出了四张古老的黑胶唱片，逐个播放起来。我从A大调的前奏曲和赋格（*BWV 536*，施韦泽取名为《行走的赋格》）开始听，经过了很多其他的乐曲，终于又听到了我最喜爱的美妙的G大调前奏曲和赋格（*BWV 541*）。然后，作为最后的压轴，我听了那首交织了甜美与苦涩的哀婉动人的赞美诗式前奏曲《凡人必有一死》（“我们都必须死去”——或者，要想与那句德语诗句的扬抑格步调形成共鸣，也可以翻译成：“凡人终有一死”）。

当我安静地坐在我的起居室里全神贯注地聆听那些柔软的深不可测的冥想音符时，我发觉一只孤单的蚂蚱正坐在地毯上。最初，我以为它是死的（毕竟，所有的蚂蚱也必有一死），可是当我靠近它的时候，它做了一次远远的跳跃。于是我迅速地从旁边的桌子上抓起了一个玻璃碗，反过来困住了那个小小的跳跃者，然后小心翼翼地从下面滑入一张唱片的封面包装，给那个玻璃房间加了块地板。然后，我把这个即兴而做的手工艺品和里面小小的乘客一起送到了我的前门，打开门，让蚂蚱跳入了黑夜中的一片灌木丛中。只有在我执行这个小小的善行的过程当中，我的脑中才真正发生了与施韦泽精神的共鸣——事实上，这是在我把那张唱片的封面滑到玻璃碗底下的时候发生的，那张封面上有本·沙恩（Ben Shahn）画的施韦泽演奏风琴的肖像，那只蚂蚱正坐在施韦泽的手上。这次相遇感觉非常荣幸。

差不多一个小时之后，我起身伸展一下四肢，碰巧注意到有一只木虱正在一张桌子下面爬行，所以我又为它制作了一个小小的运输工具，把我这位6只脚的朋友护送出门外。所有这些小小的善行都是在我沉浸于巴赫深刻的精神性和施韦泽“敬畏生命”平和的思想之中时发生的，我开始感觉到这个现象相当奇妙。

也许是为了打破这个魔咒，或者，也许是为了巩固我自己的界限，当我后来又看到一个小黑点沿着左拐右拐的某种熟悉的路线围着一盏台

灯飞来飞去的时候，我对它展开了追踪。那个小黑点落在了台灯下面的桌子上面，它是什么东西已经明白无疑了：那是一只蚊子，un moustique, una zanzara, eine Mücke ㊟。过了一会儿，那只Mücke就成为历史（我就不说细节了）。到这一点为止，我猜，我表达的蚊子可以被杀死的这个观点在本书的读者眼中可能已经无须赘述，但是我还是不得不说，对于刚刚那个血液追踪导弹的毁灭，我一点儿都不遗憾。

时近午夜，我中断了自己听音乐的过程，给我那住在加利福尼亚州的年迈病弱的母亲打了一个电话，因为我已经养成了习惯，每天晚上都要给她打一个电话，聊一聊家庭新闻，给她说一点儿开心的事。在我们简短的聊天结束后，我又回到了我的音乐，而当《D小调托卡塔与赋格“多里安人”》奏响时，我发现我的思绪飘向了一位密友那里，他深爱着这首音乐，而他的儿子最近刚刚被诊断出患有一种令人担忧的疾病。音乐继续，而所有那些关于我在乎的人的思绪也在继续，而人类生命那种珍贵而令人惶恐的脆弱性也不知怎么就自然而然地掺了进来。

最出乎意料的是，午夜过后的某个时刻，我听到后门有人敲门，（这可不是我们家经常发生的事，我向你保证！）于是我跑过去看看那是谁。结果那是我之前见过一两次的一个10多岁的少年，他在一个多月之前被父母踢出家门，现在睡在公园里。他说，那天夜里有点太冷了，问我能不能在我们家的游戏室里借宿一晚。我思考了片刻，鉴于我知道我的女儿是信任他的，所以我也答应了他的请求。

突然之间，这似乎变成了一次极其奇怪的巧合，所有那些充满人性的事，那些依赖于我对其他生命内在的镜照的事件，都正好当我聚焦在同情和宽宏的概念之上时发生了。

朋友

同情、宽宏、敬畏生命——所有这些特质都在阿尔贝特·施韦泽的身上得到了集中的体现；不仅如此，他还是一位受人尊敬的巴赫风琴演奏家——可在我心中，这绝非巧合。有人会说，施韦泽和达到他罕见标准的人都是无私无我的。我理解这个观点，而且我觉得这里面说到了一些真相，但是另一方面令人古怪的是我一直都在申辩，正如词源学上展示的那样，一个人越宽宏，那个人的自我或灵魂就越大，而不是越小！所以我要说，那些给我们留下“无我”印象的人，实际上都灵魂满满——也就是说，他们在自己的闹壳/大脑/心灵/灵魂中承载了很多其他的灵魂，而我认为这种心灵空间的共享并没有削弱他们的中心内核，反而扩大且丰富了它。正如沃尔特·惠特曼（Walt Whitman）在他的诗歌《自我之歌》（*Song of Myself*）中所写：“我很大，我包含了许多。”所有这些丰富性来自一个事实，那就是在昏暗的过去的某一点，人类的大脑越过了灵活性的临界阈限，变成了准通用性的，从而有能力内化其他人类大脑的抽象实质。这才是真正值得惊叹的事。

一天，正当我努力想为“意识”这个词的应用范围画一条个人的分界线时（尽管鲜明的分界线当然是不存在的），我忽然想到，这里最关键的因素是讨论的实体对象是否具有某种“朋友”的概念，也许这个概念只是很初级的，但这个朋友必得是你在意并且在意你的某个人。似乎很明显的是，人类的婴儿从相当早开始就掌握了这个概念的基础；似乎还很明显的是，某些种类的动物——几乎但不总是哺乳动物——对于“朋友”的概念也有相当成熟的认识。

很明显，狗能感觉到特定的人类和狗是它们的朋友，可能还有一些其他种类的动物。我不会试图穷举出哪种动物看起来有能力获得“朋友”的概念，因为这条界限还是很模糊的，还因为你也可以跟我同样轻易地列出一张自己的思维清单。但是我关于此越想就越觉得正确。所以我最终把自己引向了一个出乎意料的结论，那就是，在现实中，那个看上去是自我缩影的东西——关于“我”的认识，且仅当与这个自我存在情感联系的其他自我伴随这个自我存在时，它才会存在。简言之，有宽

容，才有自我。

这个关于意识本性的观点跟大多数心灵哲学家所持的观点是多么不同啊！他们的观点是，意识是所谓感受质（qualia）的结果，所有“更高级”的体验都以自下而上的方式从这种假想中的原初体验（比如紫色造成的视网膜震动、中音C的声音或者赤霞珠的味道）中搭建起来。与此相反，我的观点设定了一个高度的抽象概念作为阈限，意识由此开始从阴影中浮现出来。文字也许可以“体验”到血的味道的可感受质（quale），但是它对于那种可感受质是没有意识的，就如同马桶可以对不同水位变化的感受质做出反应，但是完全没有意识一样。但凡蚊子拥有足够大的大脑，能让他们有朋友，那么他们就能够对那个美妙的味道产生意识了！吁，可怜的小脑瓜蚊子从根本构造上就被剥夺了那个机会。

但是，多亏了令我们得以拥有朋友并感受到爱的足够复杂的大脑，我们享有了生而为人的荣耀，获赐了体验身边广大世界的福利；这也就是说，我们具有了意识。这实在不坏。



1. 在20世纪60年代末和70年代初，马丁·路德·金和罗伯特·肯尼迪的暗杀事件让我陷入了深深的苦闷之中，我与农场工人联合组织委员会（后改成“美国农场工人联盟”）有过两三年的紧密合作，首先是作为一个参与频繁的志愿者，后来又在几个月的时间里担任一个联合抵抗运用的组织者（首先是为了葡萄，后来是为了莴苣）。在这段时期，我与塞萨尔·查韦斯有过数面之缘，可是让我深感遗憾的是，我从来没有真正与他本人相识相交。

2. 虽然我深爱的经典法国百科全书《罗贝尔法语大词典》（*Le Petit Robert*）的第二版在刚好一个世纪之后宣称阿尔贝特·施韦泽是一名法国人！但这个地方在当时还属于德国。

3. This was [Chauchard].

4. 赋格（fugue），是复调音乐的一种创作形式，它不同于卡农的严格模仿，它的不同声部模仿第一声部的主题作为答题而相继进入，从而形成具有对比特点的旋律。赋格在

巴赫的音乐创作中发展到了极致。——译者注

5. 此处依次为法语、意大利语和德语中“一只蚊子”的意思。——译者注

结语
困境

毫不离谱！

在前面的24章里，我已经竭尽所能地表达出“我”为何物，这必然也意味着我同时也竭尽所能地说明了自我、灵魂、内在之光、第一人称视角、内在性、意向性和意识为何物。可以肯定，这的确是一项离谱的要求，但是我希望我讲的并不离谱。然而，在一部分读者看来，我的故事可能仍然是一个离谱——极其离谱！太过离谱——的故事。对于这些读者，我由衷地表示理解，因为我承认，确实还有一些令人困扰的问题依然待解。

依我所见，问题的关键在于当我们努力理解我们是什么的时候，我们人类就已经注定失败了。作为以纯物质构成的宇宙中的精神生物，我们对于自己本性的困惑将无止无休。我还历历在目地记得，少年时代的我在阅读有关大脑的书籍时，被迫在人生中第一次直面人类的大脑——特别是我自己的大脑——是一个必须服从物理法则的物理结构这一观点。虽然这在你看来也许有点奇怪，而现在的我回头看也觉得同样奇怪，但是在当时，这个发现着实令我大吃一惊。

简而言之，我们的困境如下。我们要么相信，我们的意识是某种有别于物理法则产物的东西，要么相信，它就是物理法则的产物——但是，无论做出哪种选择，我们都会被引向令人困扰的，也许甚至难以接受的结论。我在这最后几页里的目的，就是要硬碰硬地直面这一两难的

窘境。

二元论的引力和陷阱

我在第22章中讨论了二元论——这个观点认为，在由物理法则支配的物理实体之上，还存在一种名为“意识”的实体，它在宇宙中是隐形的、无法测量的、难以察觉的。这个概念跟西方传统宗教中的“灵魂”非常接近，颇具吸引力，因为它符合我们的日常经验，把世界一分为二，划分出——有生命的和无生命的——两种事物，而且它还为我们对自己的内在性或内在之光的体验给出了某种解释。我们心里都明白得很，否定内心这种东西的存在，即便不是不可能的，也会看起来极其荒诞。

二元论还展现出解释一种神秘二分法的希望；那种二分法把有生命的世界分成了两类实体：我自己和他人。没有二元论，这种介于主观的、第一人称的世界观和一个客观的、第三人称的世界之间的裂痕似乎再无别的弥合之法了。如果有一种不可分解的（意识）实质，在诞生的一刻便魔法般地分配给每一个人，每一份都带有独一无二的特色，永恒不变地定义了那个受体的同一性，而我们所说的“我”正是其中一份，那么我们便不需要再为我们是何物而继续寻找解释了（尽管它依赖的是某种不可解释之物）。

进一步说，认为我们每个人都由一种独特的非躯体化实质内在决定的观点，还意味着我们都拥有不朽的灵魂；因而，对二元论的信仰也许能减缓一些死亡的刺痛。对于某个在成长过程中浸淫于西方宗教的艺术和言语意象之中的人而言，要想象一种纤弱缥缈的光晕从某个刚死之人的身体中释放出来，向上飘远，直抵某个目不可及的天外国度，并永生不灭地生活在那里，并非难事。不管我们是笃信者还是怀疑者，这种意象都是我们西方文化遗产的一部分。正因如此，不管一个人立足于科学的信仰系统多么坚定，也很难把它全盘摒弃。

在我的妻子卡罗尔死后不久，我为她组织了一场纪念仪式，由几位亲近的朋友和家人分享对她的怀念，仪式上播放的都是对她意义重大的音乐。作为这场悲伤的仪式的收尾，我选择了谢尔盖·普罗科菲耶夫《第一小提琴协奏曲》开场乐章的最后2分30秒的一段音乐，那是一首非凡的音乐诗歌作品，我和卡罗尔都曾为之心醉神迷。我从这部协奏曲（以及它的姐妹篇，二者首尾相接，才构成了一个完整体）中选取的那段美妙而感人的篇章，不如说正是为了唤起一个飞升的灵魂意象而创作的，从头至尾都如此轻柔、敏感而精致，在它结尾处逐渐上扬的曲调中达到了极致。虽然我和卡罗尔对于宗教都完全没有信仰，可那个天真的意象在我听来却如此真实：卡罗尔最纯粹的实质离开了她尘世的肉体，向上飞升，一直向上，尽管她的灵魂最终没有飞入天国，而是仅仅飞入了我这个家伙……

在这个故事里，这个家伙接受过多年的科学训练，而且有关基于物理学的思维和精神展开过冷酷的思考，可是他时不时地也会受到传统二元论意象的影响，毕竟那种意象从小到大地陪伴了我们大多数人的成长——如果不是由我们家人灌输的，也会在更广阔的文化中耳濡目染。即便我拒绝那个观点，我还是可以落入那个诱人的意象之中。但是在我更理智的时刻，这种意象在我看来毫无道理可言，因为我实在太过了解二元论是如何引出了一长串无法回答的问题。有些问题，我在第22章中已经写出来了。这些问题表明，二元论充斥着武断和不连贯性，看起来其自己的重量就足以把自己压垮了。

非二元论的魅惑与罅隙

如果反过来，一个人相信意识（这里的“意识”不再是楷体）是物理法则的产物，那么任何“上面”的额外之物都没有立足之地了。这对于拥有科学头脑的人来说是很有吸引力的，因为它远比二元论简单得多。它摆脱了平凡的物理实体和超凡的非物理实体之间令人迷惑的二分法，而

那个围绕非物理性实体的本性而提出的绵长的问题清单也因此一笔勾销了。

另一方面，把二元论踢出局也同样令人不安，因为至少乍一看，这么做仿佛让我们无法再区分有生命和无生命的实体。而且我们关于自己内在性或内在之光的独特体验也无从解释，我们的自我和他人的自我之间的鸿沟也难以说明了。可是如果我们更仔细地审视一下这个观点就会发现它还是为这些区分留有空间。

在引言中，我写到过“自我和灵魂也从由无生命物质组成的基底上奇迹般地现身”；我猜，这个说法曾经让不止一位读者倒吸了一口凉气。“这个作者怎么可以说一个人类的大脑——宇宙中所有实体中最有生命力的东西——是‘无生命物质’呢？”这么说吧，本书的主旨之一便是，生命力的在场或缺席取决于人们在何种层级上观看一个结构。在最高的、最具集合性的层级上看，一个大脑是绝对有生命和有意识的。但是当视角逐渐降低，结构逐层下降，从大脑到皮层、从神经柱到神经元，从细胞质到蛋白质，再从肽到粒子，生命的感觉逐渐消失，直到最低的层级，生命的痕迹已经完全不见了。人们的思维可以在最高和最低的等级之间来回穿梭移动，并以这种方式任意摇摆在把大脑视为有生命之物和无生命之物的视角之间。

因此，只要不同层级的描述是有效的，非二元论的观点便可以轻易地涵盖有生命的实体。有生命的实体是那些在某个描述层级上表现出特定类型的环路模式；而当一个系统拥有通过知觉把世界过滤成不同范畴的内在能力，并精力十足地不断向抽象的层面扩大它的范畴库时，这种环路模式将不可避免地开始生成。当根深蒂固的自我表征——实体给自己讲的故事——出现时，这种模式也就完全成熟了，由实体的“我”在其中担任主角，在一系列欲望的驱动下形成一个统一的因果主体。更加精确地说，一个实体拥有生命的程度，取决于这种环路“我”的模式的存在性，因为这种模式的存在本来就不是一件全或无的事。因此，在一个给

定基底上存在一个“我”的模式的程度，决定了生命的程度，没有这种模式，则实体也没有生命。

彩虹还是岩石？

还有一个阴魂不散的问题：一个环路的抽象模式，不管它有多么精妙，到底是如何构成了一个内在性之所、一束光、一个第一人称经验的场地呢？换句话说，“我”从何而来？认为这样一种模式可以随着时间的推移不断变大、变复杂、感知自己、扎根于自己内部而最终变得不可动摇，这种说法对于某些真理的追求者而言已经是一个令人满意的答案了（比如怪圈641号）。然而，对于另外一些人（比如怪圈642号）来说，还远远不够。

在后一种思考者而言，第21章提出的那个谜题永难消散。两个根据一个被毁的身体全新制造出来的精细到原子的复制品，一个在火星上，一个在金星上：“我将在哪里苏醒？如果两个身体之一将承载我的内在之光，那么到底是哪一个？”这类思考者疯狂地抓住了一个直觉性的概念不放，认定任何一个有知生命的同一性、“我”、内在之光和内在性，都是由一个独一无二的笛卡儿自我构成的。在这些思考者看来，说他们关于“我”的珍贵概念更像是一道闪烁的、难以把握的彩虹，而不像一块坚实的岩石，因而“哪一个才是我”这个令人迷惑的谜题根本就没有正确答案，那简直是难以接受的。他们会坚称，在两个身体中的一个而非另一个中，一定有一个真正的玻璃球“我”，而不是一个难以捉摸的彩虹一样的实体：当我们逐渐靠近时，它先是后退躲避，然后整个分解不见。但是，相信这样一种不可分、不可解的“我”，就相当于相信了非物理性的二元论。

猛刺：硬的问题

下面我们要走入我们核心的困境了。我们要么就要相信有非物理性的灵魂生活在物理法则之外，这等同于抛弃科学，相信魔法，要么就要拒绝这个观念，可这样一来，“是什么让一个纯粹物理性的模式成为我？”这个挥之不去的问题——哲学家大卫·查尔默斯充满挑逗意味地把它命名为“硬的问题”（The Hard Problem）——看起来就像几个世纪以前一样（以及未来任何时候）找不到答案了。

说到底，像“物理系统”或“物理基底”这样的短语给大多数人——也包括世界上很重要的一部分哲学家和神经学家——留下的印象，都是一个由大量互相交错的轮子、齿轮、铁钉、管道、球体、摆锤等等部件构成的精密结构，尽管这些部件可能是微小的、不可见的，甚至可能是概率性的。这样一种由交互的非生命材料构成的事物在大多人看来都是无意识的，缺乏内在之光，跟抽水马桶、自动变速器、精美的瑞士手表（不管是机械的还是电子的）、齿轨铁路、远洋油轮和炼油厂一样。在他们看来，这种系统不仅仅可能是无意识的，而且必然如此。这正是约翰·塞尔以巧妙的手腕加以利用的那种单一层级的直觉，他试图以此劝人们相信，不管计算机具有何种抽象模式，它永远不可能有意识；不管它组织出多么长的词汇链条，它也永远不可能表达任何意义。

回刺：一首温柔的诗㊟

我忠实的读者，你已历尽艰辛跋涉至此，马上就到最后一页了，我希望你的眼中所见能有所不同。我和你共同游历了一个又一个精密度逐渐递增的环路结构，从探索科学博物馆那个不停飞走的红点，到拍摄自己所连接的屏幕的高分辨率的电视摄像头，再到声明自身没有*PM*证明的共识，最终抵达从每个人类大脑中不停增长的符号库里现身的怪圈。（心灵活力是我们所不能容忍的，因为它引向无穷无尽的陷阱。）

如果在我们由物理学支配的世界里真的有某种魔法，那么它必然存

在于这些自我反射、自我定义的模式当中。哥德尔的“特洛伊木马”把自我意识偷偷运入了那座专门为了抵御它而建造的坚不可摧的城堡之中，激发了怪圈的灵感，让人想起罗杰·斯佩里那座在“力中之力中之力之塔”（可以在每个于恐惧与梦想间摇摆不定的灯泡内部找到）。而这种怪圈是我能想象到的唯一解释，可以说明有生命的、欲望驱动的生命如何从平凡的物质中脱颖而出，而且还能说明，在充斥着我们这个星球的各种各样的环路模式中，有且只有一个是你称之为“我”（而我称之为“你”）的那个。

一条腿上的亿兆只蚂蚁

我和你是感知自己的幻景，这幕后唯一的魔法机制就是感知——在庞大的原始数据流的作用下，触发了一个小小的符号集合；那些符号则代表着这个世界里的抽象规则。当抽象级别相当高的感知进入物理世界，而且大量反馈环开始运作时，“它”终于变成了“他”。那些曾经被粗暴地标记为“机械的”东西，那些曾经被本能地排除在意识候选人之外的东西，就不得不被重新考虑了。

我们人类是宇宙中的宏观结构，而这个宇宙的法则在一个微观的层面上运转。作为渴望生存的生命，我们有十足的动力去寻求只利用我们自己所处层级的实体来进行解释的有效方法。因此，我们在我们容易感知的实体周围确定了概念的边界，在这么做的同时，我们也塑造了我们眼中的现实。我们为每个人创造的那个“我”是这种感知或发明的现实的一个典型范例，它如此出色地解释了我们的行为，而仿佛成为世界的中心。但是这个“我”的概念，其实是大量翻滚蒸腾的粒子，而那些翻滚蒸腾的粒子，我们必然是察觉不到的。

有时，当我的腿“睡着了”（这是英语里形容腿麻的说法），我感到里面有上千根细针在刺挠的时候，我会对我自己说：“啊哈！所以这就

是活着的真正感觉！我正在罕见地体会到真正的我有多么复杂！”(注)。当然，我们从来没有接近过体验到我们真正拥有的完整的复杂性，因为——拿一个典型的例子来说——我们拥有60亿兆（也就是60万亿亿）个血红蛋白分子，它们无时无刻不在我们的血管中乱窜，而在我们生命的每一秒中，它们中的400万亿个会被毁灭，而又有400万亿个同时被创造出来。像这样的数字远远地超出了人类的感知范围。

但是，我们自己的不可测性对我们而言却是件幸运的事！这就好比，如果我们能够真正领会我们与我们生于其中的广袤宇宙相比究竟有多么渺小，那么我们可能会萎靡至死；同理，如果我们要了解发生在我们身体内部的难以想象的疯狂世界，我们也可能会在恐惧和惊慌中手足无措。我们生活在一种受到祝福的无知状态中，但是它同时也是一种神奇的启蒙状态，因为它让我们漂浮在一个由我们自己创造的中级范畴的宇宙之中——那些范畴在促进生存的方面发挥了不可思议的功能。

我是个怪圈

在最后，我们自我感知、发明自我并被锁入一场叫作“我”的幻景，是自我指涉的小小奇迹。我们相信那些玻璃球，虽然当我们搜寻它们的时候，它们会消解不在，但是当我们不去寻找它们时，它们却和任何一颗真正的玻璃球一样真实。我们最真的本性就这样阻止我们完全理解它最真的本性。我们人类静静地安处于非可视化的宇宙尺度的曲线时空与含糊不清、影影绰绰的带电粒子之间的某个中间地带，更像彩虹和海市蜃楼，而不是雨滴或岩石。我们是我们自己创作的不可预知的诗歌——暧昧、隐喻、含糊，有时呈现超乎寻常的美丽。

用这种方式看待我们自己，可能不如信仰某种妙不可言的永恒存在的彼岸那样令人宽慰，但它也有足以补偿的好处。我们放弃的是一种幼稚的认识，认为事物就是它们显现出来的样子，而我们那个看起来很坚

固、像玻璃球一样的“我”就是世界上最真实之物；而我们的收获是欣赏到了我们在我们的核心处是多么精妙，我们跟我们表面上看起来相比有多么不同。库尔特·哥德尔用他那个出乎意料的怪圈为我们打开了一个观察数学为何物的更深且更微妙的视野；同样，这里用怪圈描述我们的实质，也为我们打开了一个更深的且更微妙的观察何为人类的视野。在我的心里，这是利大于弊的。

---

1. 在这一部分，我的疯狂之中是有条理可循的。特别是，两个自然段都是用一种名为“四音节”的古老音步写成的。它的意思是，连续三个音节没有重音，而在第四个音节处加入一个重音，但是（我希望）不要显得太过牵强：“And yet to *you*, my faithful *reader* who has *plowed* all through this *book* up to its nearly final *page*…”每个自然段的最后一个辅音都在“音步”的意义上恰切到位（意味着是重读的引介）。这些“音节”的数量必须是40个，原因在于，我是在模仿*Le Ton beau*的第5a页。

2. 在法语里，人们会说一个人“腿里有蚂蚁”，而卡通人物淘气阿丹（Dennis the Menace）曾经形容说，他感觉“腿里有姜汁汽水”。这是两种对这一古怪而又普遍的感觉的令人难忘的比喻。（我在这个脚注中的疯狂是有条理可循的。特别是，这个脚注既描述了又呈现了一种名为“四音节”的古老音步。它的意思是，连续三个音节没有重音，而在第四个音节处加入一个重音，但是（我希望）不要显得太过牵强。下面，我为您提供一个简短的例子，并建议您把它大声读出来：“There is a method to my madness in this footnote…”（我在这个脚注中的疯狂是有条理可循的）。特别是，我使用了正正好好40个音步，因为我是在模仿*I Am a Strange Loop*的第367页上的由40个音步构成的两个段落。）

参考文献

Aitchison, Jean. *Words in the Mind: An Introduction to the Mental Lexicon* (second edition). Cambridge, Mass.: Blackwell, 1994.

Andersen, Peter B. *et al.* (eds.). *Downward Causation: Minds, Bodies, and Matter*. Aarhus: Aarhus University Press, 2000.

Anderson, Alan Ross. *Minds and Machines*. Englewood Cliffs, N.J.: Prentice-Hall, 1964.

Applewhite, Philip B. *Molecular Gods: How Molecules Determine Our Behavior*. Englewood Cliffs, New Jersey: Prentice-Hall, 1981.

Ash, Avner and Robert Gross. *Fearless Symmetry: Exposing the Hidden Patterns of Numbers*. Princeton: Princeton University Press, 2006.

Atlan, Henri. *Entre le cristal et la fumée: Essai sur l'organisation du vivant*. Paris: Éditions du Seuil, 1979.

Barwise, K. Jon and Lawrence S. Moss. *Vicious Circles: On the Mathematics of Non-wellfounded Phenomena*. Cambridge, U.K.: Cambridge University Press, 1996.

Berg, Paul and Maxine Singer. *Dealing with Genes: The Language of Heredity*. Mill Valley, Calif.: University Science Books, 1992.

Bewersdorff, Jörg. *Galois Theory for Beginners*. Providence: Am. Mathematical Society, 2006.

Bierce, Ambrose. “An Occurrence at Owl Creek Bridge”. In *The Collected Writings of Ambrose Bierce*. New York: Citadel Press, 1946.

Blackmore, Susan. *The Meme Machine*. New York: Oxford University Press, 1999.

Boden, Margaret A. *The Creative Mind: Myths and Mechanisms*. New York: Basic Books, 1990.

Boolos, George S. and Richard C. Jeffrey. *Computability and Logic*. New York: Cambridge University Press, 1974.

Borges, Jorge Luis. *Ficciones*. New York: Grove Press, 1962.

Bougnoux, Daniel. *Vices et vertus des cercles: L’autoréférence en poétique et pragmatique*. Paris: Éditions La Découverte, 1989.

Braitenberg, Valentino. *Vehicles: Experiments in Synthetic Psychology*. Cambridge, Mass.: MIT Press, 1984.

Brinck, Ingar. *The Indexical “I”: The First Person in Thought and Language*. Dordrecht: Kluwer, 1997.

Brown, James Robert. *Philosophy of Mathematics*. New York: Routledge, 1999.

Carnap, Rudolf. *The Logical Syntax of Language*. Paterson, N.J.: Littlefield, Adams, 1959.

Casati, Roberto and Achille Varzi. *Holes and Other Superficialities*. Cambridge, Mass.: MIT Press, 1994.

—. *Unsurmountable Simplicities: Thirty-nine Philosophical*

Conundrums. New York: Columbia University Press, 2006.

Chaitin, Gregory J. *Information, Randomness, and Incompleteness: Papers on Algorithmic Information Theory*. Singapore: World Scientific, 1987.

Chalmers, David J. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press, 1996.

Chauchard, Paul. *Le Cerveau et la conscience*. Paris: Éditions du Seuil, 1960.

Chandrasekhar, B. S. *Why Things Are the Way They Are*. New York: Cambridge University Press, 1998.

Churchland, Patricia. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, Mass.: MIT Press, 1986.

Cope, David. *Virtual Music: Computer Synthesis of Musical Style*. Cambridge, Mass.: MIT Press, 2001.

Cordeschi, Roberto. *The Discovery of the Artificial: Behavior, Mind, and Machines Before and Beyond Cybernetics*. Dordrecht: Kluwer, 2002.

Dahl, Roald. *Kiss Kiss*. New York: Alfred A. Knopf, 1959.

Damasio, Antonio. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace, 1999.

Davies, Philip J. “Are there coincidences in mathematics?” *American Mathematical Monthly* 88 (1981), pp. 311–320.

Davis, Joshua. “Say Hello to Stanley”. *Wired* 14 (January 2006).

Davis, Martin (ed.). *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems, and Computable Functions*. Hewlett, N.Y.: Raven, 1965.

Dawkins, Richard. *The Selfish Gene*. New York: Oxford University Press, 1976.

DeLong, Howard. *A Profile of Mathematical Logic*. Reading, Mass.: Addison-Wesley, 1970. (Reissued by Dover Press, 2004.)

Dennett, Daniel C. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, Mass.: MIT Press, 1978.

—. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: MIT Press, 1984.

—. *The Intentional Stance*. Cambridge, Mass.: MIT Press, 1987.

—. *Consciousness Explained*. Boston: Little, Brown, 1991.

—. "The Self as a Center of Narrative Gravity", in F. Kessel, P. Cole, and D. Johnson (eds.), *Self and Consciousness*. Hillsdale, N.J.: Lawrence Erlbaum, 1992.

—. *Kinds of Minds: Toward an Understanding of Consciousness*. New York, Basic, 1996.

—. *Brainchildren: Essays on Designing Minds*. Cambridge, Mass.: MIT Press, 1998.

—. *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, Mass.: MIT Press, 2005.

Donald, Merlin. *A Mind So Rare: The Evolution of Human Consciousness*. New York: W. W. Norton, 2001.

Dupuy, Jean-Pierre. *Ordres et Désordres*. Paris: Éditions du Seuil, 1982.

—. *The Mechanization of the Mind: On the Origins of Cognitive Science*. Princeton: Princeton University Press, 2000.

Edson, Russell. *The Clam Theater*. Middletown, Conn.: Wesleyan University Press, 1973.

Enrustle, Y. Ted. *Prince Hyppia: Math Dramatica*, Volumes I–III. Luna City: Unlimited Books, Ltd., 1910–1913.

Falletta, Nicholas. *The Paradoxicon*. New York: John Wiley & Sons, 1983.

Fauconnier, Gilles. *Mental Spaces*. Cambridge, Mass.: MIT Press, 1985.

Flanagan, Owen. *The Science of the Mind*. Cambridge, Mass.: MIT Press, 1984.

Gamow, George. *One Two Three… Infinity*. New York: Mentor, 1953.

Gardner, Martin. *Wheels, Life, and Other Mathematical Amusements*. New York: W. H. Freeman, 1983.

Gebstadter, Egbert B. *U Are an Odd Ball*. Perth: Acidic Books, 2007.

Gentner, Dedre, Keith J. Holyoak, and Boicho N. Kokinov (eds.). *The Analogical Mind: Perspectives from Cognitive Science*. Cambridge, Mass.: MIT Press, 2001.

Gödel, Kurt. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. New York: Basic Books, 1962. (Reissued by Dover, 1992.)

Goldstein, Rebecca. *Incompleteness: The Proof and Paradox of Kurt Gödel*. New York: W. W. Norton, 2005.

Goodstein, R. L. *Development of Mathematical Logic*. New York: Springer, 1971.

Greenstein, George and Arthur G. Zajonc. *The Quantum Challenge*. Sudbury, Mass.: Jones and Bartlett, 1997.

Griffin, Donald R. *The Question of Animal Awareness*. New York: Rockefeller U. Press, 1976.

Hardy, G. H. and E. M. Wright. *An Introduction to the Theory of Numbers*. New York: Oxford University Press, 1960.

Hart, Leslie A. *How the Brain Works*. New York: Basic Books, 1975.

Harth, Erich. *Windows on the Mind: Reflections on the Physical Basis of Consciousness*. New York: William Morrow, 1982.

Haugeland, John (ed.). *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Montgomery, Vermont: Bradford Books, 1981.

—. *Artificial Intelligence: The Very Idea*. Cambridge, Mass.: MIT Press, 1985.

Hennie, Fred. *Introduction to Computability*. Reading, Mass.: Addison-Wesley, 1977.

Hoffmann, Banesh. *Albert Einstein, Creator and Rebel*. New York: Viking, 1972.

Hofstadter, Douglas R. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books, 1979. (Twentieth-anniversary edition published in 1999.)

—. “Analogies and Metaphors to Explain Gödel’s Theorem”. *The Two-Year College Mathematics Journal*, Vol. 13, No. 2 (March 1982), pp. 98–114.

—. *Metamagical Themas: Questing for the Essence of Mind and Pattern*. New York: Basic Books, 1985.

—. *Le Ton beau de Marot: In Praise of the Music of Language*. New York: Basic Books, 1997.

—. “Analogy as the Core of Cognition”. Epilogue to D. Gentner, K. Holyoak, and B. Kokinov (eds.), *The Analogical Mind*. Cambridge, Mass.: MIT Press, 2001.

—. “Mystery, Classicism, Elegance: an Endless Chase after Magic”. In D. Schattschneider and M. Emmer (eds.), *M. C. Escher’s Legacy*. New York: Springer, 2002.

Hofstadter, Douglas R. and Daniel C. Dennett (eds.). *The Mind’s I: Fantasies and Reflections on Self and Soul*. New York: Basic Books, 1981.

Hofstadter, Douglas R. and David J. Moser. “To Err Is Human; To Study Error-making Is Cognitive Science”. *Michigan Quarterly Review* 28, no. 2 (1989), pp. 185–215.

Hofstadter, Douglas R. and the Fluid Analogies Research Group. *Fluid Concepts and Creative Analogies*. New York: Basic Books, 1995.

Holland, John. *Hidden Order: How Adaptation Builds Complexity*. Redwood City, Calif.: Addison-Wesley, 1995.

—. *Emergence: From Chaos to Order*. Redwood City, Calif.: Addison-Wesley, 1997.

Horney, Karen. *Self-Analysis*. New York: W. W. Norton, 1942.

—. *Our Inner Conflicts: A Constructive Theory of Neurosis*. New York: W. W. Norton, 1945.

Hughes, Patrick and George Brecht. *Vicious Circles and Paradoxes*. New York: Doubleday, 1975.

Huneker, James. *Chopin: The Man and His Music*. New York: Scribner's, 1921. (Reissued by Dover, 1966.)

Huntley, H. E. *The Divine Proportion: A Study in Mathematical Beauty*. New York: Dover, 1970.

Jauch, J. M. *Are Quanta Real? A Galilean Dialogue*. Bloomington: Indiana University Press, 1989.

Jeffrey, Richard C. *Formal Logic: Its Scope and Limits*. New York: McGraw-Hill, 1967.

Judson, Horace Freeland. *The Eighth Day of Creation*. New York: Simon & Schuster, 1979.

Kahneman, Daniel and Dale Miller. "Norm Theory: Comparing Reality

to Its Alternatives”. *Psychological Review* 80 (1986), pp. 136–153.

Kanerva, Pentti. *Sparse Distributed Memory*. Cambridge, Mass.: MIT Press, 1988.

Kent, Jack. *Mr. Meebles*. New York: Parents’ Magazine Press, 1970.

Klagsbrun, Francine. *Married People: Staying Together in the Age of Divorce*. New York: Bantam, 1985.

Kneebone, G. T. *Mathematical Logic and the Foundations of Mathematics*. New York: Van Nostrand, 1963.

Kolak, Daniel. *I Am You: The Metaphysical Foundations for Global Ethics*. Norwell, Mass.: Springer, 2004.

Kriegel, Uriah and Kenneth Williford (eds.). *Self-Representational Approaches to Consciousness*. Cambridge, Mass.: MIT Press, 2006.

Kuffler, Stephen W. and John G. Nicholls. *From Neuron to Brain*. Sunderland, Mass.: Sinauer Associates, 1976.

Külot, Gerd. “On Formerly Unpennable Proclamations in *Prince Hyppia: Math Dramatica* and Related Stageplays (I)”. *Bologna Literary Review of Bologna* 641 (1931).

Laughlin, Robert B. *A Different Universe: Reinventing Physics from the Bottom Down*. New York: Basic Books, 2005.

Le Lionnais, François. *Les Nombres remarquables*. Paris: Hermann, 1983.

Lem, Stanislaw. *The Cyberiad: Fables for the Cybernetic Age*

(translated by Michael Kandel). San Diego: Harcourt Brace, 1985.

Livio, Mario. *The Equation that Couldn’t Be Solved*. New York: Simon and Schuster, 2005.

Margolis, Howard. *Patterns, Thinking, and Cognition*. Chicago: University of Chicago, 1987.

Martin, Richard M. *Truth and Denotation: A Study in Semantical Theory*. Chicago: University of Chicago Press, 1958.

McCorduck, Pamela. *Machines Who Think*. San Francisco: W. H. Freeman, 1979.

Mettrie, Julien Offray de la. *Man a Machine*. La Salle, Illinois: Open Court, 1912.

Metzinger, Thomas. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, Mass.: MIT Press, 2003.

Miller, Fred D. and Nicholas D. Smith. *Thought Probes: Philosophy through Science Fiction*. Englewood Cliffs: Prentice-Hall, 1981.

Minsky, Marvin. *The Society of Mind*. New York: Simon & Schuster, 1986.

—. *The Emotion Machine*. New York: Simon & Schuster, 2006.

Monod, Jacques. *Chance and Necessity*. New York: Vintage Press, 1972.

Moravec, Hans. *Robot: Mere Machine to Transcendent Mind*. New York: Oxford University Press, 1999.

Morden, Michael. “Free will, self-causation, and strange loops”. *Australasian Journal of Philosophy* 68 (1990), pp. 59–73.

Nagel, Ernest and James R. Newman. *Gödel’s Proof*. New York: New York University Press, 1958. (Revised edition, edited by Douglas R. Hofstadter, 2001.)

Neumann, John von. *Theory of Self-Reproducing Automata* (edited and completed by Arthur W. Burks). Urbana: University of Illinois Press, 1966.

Niven, Ivan and Herbert S. Zuckerman. *An Introduction to the Theory of Numbers*. New York: John Wiley & Sons, 1960.

Nørretranders, Tor. *The User Illusion*. New York: Viking, 1998.

Nozick, Robert. *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press, 1981.

Pais, Abraham. *Inward Bound: Of Matter and Forces in the Physical World*. New York: Oxford University Press, 1986.

—. *Niels Bohr’s Times*. New York: Oxford University Press, 1991.

Parfit, Derek. *Reasons and Persons*. New York: Oxford University Press, 1984.

Pattee, Howard H. *Hierarchy Theory: The Challenge of Complex Systems*. New York: Braziller, 1973.

Peitgen, H.-O. and P. H. Richter. *The Beauty of Fractals*. New York: Springer, 1986.

Penfield, Wilder and Lamar Roberts. *Speech and Brain-Mechanisms*.

Princeton: Princeton University Press, 1959.

Penrose, Roger. *The Emperor's New Mind*. New York: Oxford University Press, 1989.

Perry, John (ed.). *Personal Identity*. Berkeley: University of California Press, 1975.

Péter, Rózsa. *Recursive Functions*. New York: Academic Press, 1967.

Pfeiffer, John. *The Human Brain*. New York: Harper Bros., 1961.

Poundstone, William. *The Recursive Universe*. New York: William Morrow, 1984.

Pullman, Bernard. *The Atom in the History of Human Thought*. New York: Oxford University Press, 1998.

Pushkin, Alexander S. *Eugene Onegin: A Novel in Verse* (translated by James Falen). New York: Oxford University Press, 1995.

—. *Eugene Onegin: A Novel Versification* (translated by Douglas Hofstadter). New York: Basic Books, 1999.

Quine, Willard Van Orman. *The Ways of Paradox, and Other Essays*. Cambridge, Mass.: Harvard University Press, 1976.

Ringle, Martin. *Philosophical Perspectives in Artificial Intelligence*. Atlantic Highlands: Humanities Press, 1979.

Rucker, Rudy. *Infinity and the Mind*. Boston: Birkhäuser, 1982.

Sander, Emmanuel. *L'analogie, du Naïf au Créatif: Analogie et*

Catégorisation. Paris: Éditions L'Harmattan, 2000.

Schank, Roger C. *Dynamic Memory*. New York: Cambridge University Press, 1982.

Schweitzer, Albert. *Aus Meiner Kindheit und Jugendzeit*. Munich: C. H. Beck, 1924.

Searle, John. "The Myth of the Computer" (review of *The Mind's I*). *The New York Review of Books*, April 29, 1982, pp. 3–6.

Shanker, S. G. (ed.). *Gödel's Theorem in Focus*. New York: Routledge, 1988.

Simon, Herbert A. *The Sciences of the Artificial*. Cambridge, Mass.: MIT Press, 1969.

Singer, Peter and Jim Mason. *The Way We Eat: Why Our Food Choices Matter*. Emmaus, Pennsylvania: Rodale Press, 2006.

Skinner, B. F. *About Behaviorism*. New York: Random House, 1974.

Smullyan, Raymond M. *Theory of Formal Systems*. Princeton: Princeton Univ. Press, 1961.

—. *The Tao Is Silent*. New York: Harper & Row, 1977.

—. *What Is the Name of This Book?* Englewood Cliffs, New Jersey: Prentice-Hall, 1978.

—. *This Book Needs No Title*. Englewood Cliffs, New Jersey: Prentice-Hall, 1980.

—. *5000 B.C. and Other Philosophical Fantasies*. New York: St. Martin’s Press, 1983.

—. *Gödel’s Incompleteness Theorems*. New York: Oxford University Press, 1992.

Sperry, Roger. “Mind, Brain, and Humanist Values”, in John R. Platt (ed.), *New Views on the Nature of Man*. Chicago: University of Chicago Press, 1965.

Steiner, George. *After Babel*. New York: Oxford University Press, 1975.

Stewart, Ian. *Galois Theory* (second edition). New York: Chapman and Hall, 1989.

Suppes, Patrick C. *Introduction to Logic*. New York: Van Nostrand, 1957.

Thigpen, Corbett H. and Hervey M. Cleckley. *The Three Faces of Eve*. New York: McGraw-Hill, 1957.

Treisman, Anne. “Features and Objects: The Fourteenth Bartlett Memorial Lecture”. *Cognitive Psychology* 12, no. 12 (1980), pp. 97–136.

Ulam, Stanislaw. *Adventures of a Mathematician*. New York: Scribner’s, 1976.

Unger, Peter. “Why There Are No People”. *Midwest Studies in Philosophy*, 4 (1979).

—. “I Do Not Exist”. In G. F. MacDonald (ed.), *Perception and Identity*. Ithaca: Cornell University Press, 1979.

Wadhead, Rosalyn. *The Posh Shop Picketeers*. Tananarive: Wowser & Genius, 1931.

Webb, Judson. *Mechanism, Mentalism, and Metamathematics*. Boston: D. Reidel, 1980.

Weinberg, Steven. *Dreams of a Final Theory*. New York: Pantheon, 1992.

—. *Facing Up*. Cambridge, Mass.: Harvard University Press, 2001.

Wells, David G. *The Penguin Dictionary of Curious and Interesting Numbers*. New York: Viking Penguin, 1986.

—. *Prime Numbers*. New York: John Wiley & Sons, 2005.

Wheelis, Allen. *The Quest for Identity*. New York: W. W. Norton, 1958.

Whitehead, Alfred North and Bertrand Russell. *Principia Mathematica*, Volumes I–III. London: Cambridge University Press, 1910–1913.

Wilder, Raymond L. *Introduction to the Foundations of Mathematics*. New York: John Wiley & Sons, 1952.

Wolf, Robert S. *A Tour through Mathematical Logic*. Washington, D.C.: The Mathematical Association of America, 2005.

Wooldridge, Dean. *Mechanical Man: The Physical Basis of Intelligent Life*. New York: McGraw- Hill, 1968.

Wynne, Clive D. L. *Do Animals Think?* Princeton: Princeton University Press, 2004.

Yourgrau, Palle. *A World Without Time: The Forgotten Legacy of Gödel and Einstein*. New York: Basic Books, 2005.

译后记

阅读侯世达先生的这本书，不仅是智力上的挑战，更是一次对自己心灵的拷问。这部成熟、完整而开放的作品，层次丰富、结构缜密、环环相扣、首尾贯穿，又指向一个更高的终极。如果让我为这部著作找一个类比物，我觉得最恰当的就是诗歌：格律明晰、意义繁复、字字珠玑、有无限想象的可能，在磨炼思维的同时也带给人格外的享受。

但这仅限于阅读。翻译这本书，在汉语的语境中还原这本书的全部效果和价值，也跟翻译诗歌一样，几乎是不可能完成的任务。且不说作者的博学与机智、书写的厚重而轻盈或者用词上的格外考究，单是书中为主题服务的各种以英语为载体的语言游戏，就很难于汉语中再现。

最让我耿耿于怀的，是在书中地位非常重要的一对名词，simm和simmball。simm是由四个英文单词首字母缩写构成的新词，意思是“具有磁性的交互冲撞的小弹珠”，所以我按照同样的规则，把它译为“小磁交弹”，而simmball加了一个ball（球），所以我把它译成“小磁交弹球”。这样翻译虽然忠实原意，但显得有点愚蠢，最要命的问题在于，simmball其实是symbol（符号）的谐音，同时也是符号的类比，这种意义和音形同时对应的关系，在翻译成中文之后，便看不出来了。而如果不翻译simm和simmball，不了解英文的读者也很难一眼看出这两个本不存在的英文单词有什么含义，也无法了解它们的构成逻辑。于是，这一组在阅读过程中让人惊叹的巧妙设计，在翻译过程中便成为一道难以逾越的鸿沟。

此类的现象在书中不止一例，而我在接受了不得两全的前提下，始终按照以表意为先的准则进行了翻译，并尽可能地把所有英语文字游戏

转换到汉语的语境之中。比如把Epi翻译成“副”（第7章），把“I eeeoai o ooa a e ooieee o oe.”翻译成“文口舌殳月扁方而下口舌殳月音自。”（第4章）

除了这些特别的语言游戏外，我对于全书文字的处理，努力贯彻“直译”的策略，争取做到一个词对应一个词，一句话对应一句话，在能力所及的范围内尽量保持作者行文的独特风格。不管是嬉笑怒骂，还是故作艰深，我都没有出于“照顾”读者的理解或者感受的考虑而淡化或软化原意。因为我对于这本书的阅读感受是：作者的每一个词、每一次表达、每一种语气，都带有明确的结构性目的，牵一发而动全身。当作者的能力全方位超过译者的时候，译者最明智的做法也许就是被牵着鼻子走。而对于这本书的翻译而言，这或许不失为一种恰当的选择。

翻译侯世达先生的书，是注定会失败的。我很荣幸能拥有这次失败的机会，也期待更多的人能够借着我抛出的这块砖头，不断参与到这本书的引介和解读的过程当中。很感谢湖岸出版的景雁老师把这本书的翻译交付给我，信任我、催促我，并给我鼓励和帮助，也很感谢程娲和文青的引介。本书的编辑李天济一字一句地与我讨论译稿，细心地发现并挽救了很多致命的问题。虽然合我们二人之力也未必是这本书的对手，但是编辑和译者数次为了一个词的译法而争论不休，也是值得纪念的一种经历。

我要感谢清华大学的蒋云鹏老师和北京大学的杨维宇博士在哲学专业方面为我提供的帮助，也要感谢我的朋友刘岳、魏宇、杨帆和张梦龙一直以来对我的支持。感谢周旻同学陪伴我每一次的翻译历程。感谢侯世达先生的这本书，让我意识到家人和爱人对于我的真正意义，重新梳理了回忆和现实的关系，并且不再害怕失去。最后，感谢“我”。

修佳明
北京大学万柳公寓